Kongeriget Danmark

Patent application No.: PA 2002 01792

Date of filing: 19 November 2002

Applicant:
(Name and address)

H:S Rigshospitalet
Blegdamsvej 9
DK-2100 København Ø
Denmark

Title: Methods and kits for diagnosing and treating B-Cell chronic lymphocytic leukemia (B-CLL).

IPC: -

This is to certify that the attached documents are exact copies of the above mentioned patent application as originally filed.

Patent- og Varemærkestyrelsen
Økonomi- og Erhvervsministeriet

16 December 2003

Pia Petersen

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

PATENT- OG VAREMÆRKESTYRELSEN

Methods and kits for diagnosing and treating B-Cell Chronic Lymphocytic Leukemia (B-CLL)

All patent and non-patent references cited in the present application, are hereby incorporated by reference in their entirety.

Field of invention

The present invention relates to methods and kits for detecting a particular polynucleotide sequence found to be indicative of a poor prognosis of B-CLL. This polynucleotide encodes a novel protein which in one preferred embodiment can be used as a cytokine, preferably as an interleukin. Also provided are methods for identifying further polynucleotide sequences encoding further novel proteins with similar function. Furthermore the invention relates to methods and compositions for treating B-CLL in particular poor prognosis B-CLL.

Background of invention

B-CLL is the most common form of leukaemia in Denmark, with more than 250 new cases diagnosed every year. The disease results in accumulation of CD19+CD5+CD23+ lymphocytes in the blood, bone marrow and organs of the patients. B-CLL cells are long-lived, non-dividing and locked in the G_1 phase of the cell cycle. At this time it is unknown how or why B-CLL occurs and no cure is known for B-CLL. The application of more aggressive treatment strategies has been hampered by the inability to identify reproducible and reliable prognostic predictors in patients with poor outcome in this disease. In many patients the diagnosis does not affect morbidity or mortality. Other patients suffer from an incurable cancer that inevitably results in death, regardless of treatment. Until recently this latter group of patients could not be identified at the time of diagnosis. Recently, two studies established the mutational status of immunoglobulin variable region of the heavy chain (Ig V_H) genes in B-CLL as independent prognostic markers, within each clinical stage (Damle, R.N., T. Wasil, F. Fais, F. Ghiotto, A. Valetto, S.L. Allen, A. Buchbinder, D. Budman, K. Dittmar, J. Kolitz, S.M. Lichtman, P. Schulman, V.P. Vinciguerra, K.R. Rai, M. Ferrarini, and N. Chiorazzi. 1999. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic

leukemia. *Blood* 94, no. 6:1840. Hamblin, T.J., Z. Davis, A. Gardiner, D.G. Oscier, and F.K. Stevenson. 1999. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* 94; no. 6:1848). Patients without somatic hypermutation show much shorter survival than patients with somatic hypermutation. FISH-studies of cytogenetic aberrations in B-CLL established specific abnormalities on chromosomes 11 (ATM), 12 (?), 13 (Leu-1 and-2) and 17 (p53) as independent prognostic markers, within each clinical stage (Dohner, H., S. Stilgenbauer, A. Benner, E. Leupolt, A. Krober, L. Bullinger, K. Dohner, M. Bentz, and P. Lichter. 2000. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med* 343, no. 26:1910). Very recent studies have demonstrated that independent risk prediction, using a combined analysis of Ig V_H gene mutational analysis and cytogenetics, can identify subgroups of B-CLL with median survivals ranging from less than 2.5 years to more than 15 years (Krober, A., T. Seiler, A. Benner, L. Bullinger, E. Bruckle, P. Lichter, H. Dohner, and S. Stilgenbauer. 2002. V(H) mutation status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukemia. *Blood* 100, no. 4:1410; Lin, K., P.D. Sherrington, M. Dennis, Z. Matrai, J.C. Cawley, and A.R. Pettitt. 2002. Relationship between p53 dysfunction, CD38 expression, and IgV(H) mutation in chronic lymphocytic leukemia. *Blood* 100, no. 4:1404; Oscier, D.G., A.C. Gardiner, S.J. Mould, S. Glide, Z.A. Davis, R.E. Ibbotson, M.M. Corcoran, R.M. Chapman, P.W. Thomas, J.A. Copplestone, J.A. Orchard, and T.J. Hamblin. 2002, Multivariate analysis of prognostic factors in CLL: clinical stage, IGVH gene mutational status, and loss or mutation of the p53 gene are independent prognostic factors. *Blood* 100, no. 4:1177) (see Figure 1).

It is an object of the present invention to provide an explanation of the clinical heterogeneity seen in B-CLL disease subgroups. A further object is to provide differentially expressed genes, which can be used as prognostic markers of disease and information about the differences in etiology between the two groups of B-CLL patients. Since the hitherto used process of characterising Ig VH gene mutational status of an individual patient is cumbersome, an additional goal was to find a genetic marker that can be used in an easy assay to distinguish between the two subgroups. A further object of the present invention is to provide a cure and/or treatment of B-CLL, in particular of poor prognosis B-CLL.

Summary of invention

In a first aspect the invention relates to a method for diagnosing a subtype of B-cell chronic lymphocytic leukaemia (B-CLL), said method comprising the steps of determining the presence or absence of a transcriptional or translational product of SEQ ID No 1 in a biological sample isolated from a subject. The nucleic acid sequence of SEQ ID No. 1 is set forth in Figure 8. The gene is called AMB-1 in the following. SEQ ID No 1 is a 20,000 nucleotide long sequence which provides two transcriptional products in B-CLL cells in patients with poor prognosis B-CLL. Each of the two transcriptional products consists of two exons separated by the same intron. The long mRNA sequence (SEQ ID No 4) starts at base No. 49101 of SEQ ID No 1 and the short mRNA sequence (SEQ ID No 2) starts at base No. 51417 of SEQ ID No 1. Both mRNA sequences encode an open reading frame encoding a 121 amino acid peptide (SEQ ID No 3).

As evidenced by the appended examples, the present inventors have determined that an expression product is only present in one subtype of B-CLL. A transcriptional or translational product of SEQ ID No 1 has not been found in any of the other tissue types tested (see e.g. Figure 11). Therefore there is strong evidence that a transcriptional or translational product of SEQ ID No 1 has great diagnostic value and independent prognostic value.

The vast majority of patients which show expression of the AMB-1 gene show unmutated Ig V(H) genes which is consistent with poor prognosis B-CLL. The presence of a transcriptional or translational product of the AMB-1 gene can be determined easily using standard laboratory procedures and equipment. Therefore the diagnostic method provided by the present inventors provides an easy method of diagnosis as compared to the determination of the mutation status of Ig V(H) genes.

At present the inventors believe that the B-CLL subtype characterised by the presence of a translational or transcriptional produce of SEQ ID No 1 is an independent B-CLL sub-type.

Accordingly, in a further aspect of the present invention there is provided a method for determining the stage/progress of B-CLL comprising determining the amount of a transcriptional or translational product of SEQ ID No 1 in a biological sample isolated from a subject. This aspect is supported by the finding of a transcriptional product of SEQ ID No 1 in B-CLL cells. The method may be used e.g. for determining the efficiency of a treatment, i.e. to see whether the amount of the transcriptional or translational product decreases or increases in response to a curative treatment.

In a further aspect the invention relates to a method of treating B-CLL comprising administering to a subject being diagnosed according to the invention, a therapeutically effective amount of a compound capable of selectively killing and/or inhibiting division of and/or inducing apoptosis in B-CLL cells. The compound may be selected from the group chemotherapeutic agents, anti CD20, anti-CD-52, or other antibodies. The treatment may comprise using non-myelooablative bone marrow transplantation. This aspect is based on the identification of a novel sub-type of B-CLL characterised by the presence of a transcriptional or translational product of SEQ ID No 1.

In a further therapeutic aspect the invention relates to a method for treating B-CLL comprising administering to a subject with a B-CLL diagnosis a compound capable of decreasing or inhibiting the formation of a transcriptional and/or translational product from SEQ ID No 1. The present inventors believe that the presence of said transcriptional or translational product is an etiological factor in B-CLL and that the disease can be treated or cured by inhibiting the expression of such product and/or by inhibiting the effect of such product by e.g. rendering it inactive.

In one aspect the invention relates to a gene therapy vector capable of inhibiting or decreasing the formation of a transcriptional or translational product of SEQ ID No. 1. This gene therapy vector can be used for treating B-CLL based on the finding that the AMB-1 gene encoded by SEQ ID No 1 is a etiological factor in B-CLL.

The invention also relates to a novel class of proteins. These may be described a group of isolated polypeptides comprising or essentially consisting of the amino acid sequence of SEQ ID No. 3, or a fragment thereof, or a polypeptide functionally

equivalent to SEQ ID No. 3, or a fragment thereof, wherein said fragment or functionally equivalent polypeptide has at least 60% sequence identity with the polypeptide of SEQ ID No 3, and

a) has interleukin or cytokine activity; and/or

b) is recognised by an antibody, or a binding fragment thereof, which is capable of recognising an epitope, wherein said epitope is comprised within a polypeptide having the amino acid sequence of SEQ ID No 3; and/or

c) is competing with a polypeptide having the amino acid sequence as shown in SEQ ID No 3 for binding to at least one predetermined binding partner.

The protein encoded by SEQ ID No 1, the sequence of which is set forth in SEQ ID No 3 shares a very small sequence identity with any known protein. However it has been possible to use 2D and 3D analytical tools to identify the protein as a 4-helical cytokine. The 3D structure of the protein is very similar to 4-helical cytokines and in particular to IL4.

IL4 is an important cytokine in B-CLL biology. IL4 is not expressed by B-CLL cells, but the IL4 receptor is found on the cells. The IL4 that stimulates B-CLL cells is believed to be produced by T-lymphocytes. The role of IL4 in B-CLL biology is complicated. It has been suggested that IL4 can inhibit B-CLL DNA synthesis and proliferation (Luo, H.Y., M. Rubio, G. Biron, G. Delespesse, and M. Sarfati. 1991. Antiproliferative effect of interleukin-4 in B chronic lymphocytic leukemia. *J Immunother* 10, no. 6:418). Other reports demonstrated that IL4 protects B-CLL cells from apoptosis by upregulating Bcl-2 (Dancescu, M., M. Rubio-Trujillo, G. Biron, D. Bron, G. Delespesse, and M. Sarfati. 1992. Interleukin 4 protects chronic lymphocytic leukemic B cells from death by apoptosis and upregulates Bcl-2 expression. *J Exp Med* 176, no. 5:1319), and IL4 was shown to inhibit apoptosis without stimulating proliferation (Panayiotidis, P., K. Ganeshaguru, S.A. Jabbar, and A.V. Hoffbrand. 1993. Interleukin-4 inhibits apoptotic cell death and loss of the bcl-2 protein in B-chronic lymphocytic leukaemia cells in vitro. *Br J Haematol* 85, no. 3:439). Recently, a clinical study in Sweden has confirmed these in vitro studies since IL4 administration to B-CLL patients resulted in increased numbers of B-CLL cells in the blood, suggesting that IL4 had a stimulatory or antiapoptotic effect on the B-CLL cells in vivo (Lundin, J., E. Kimby, L. Bergmann, T. Karakas, H. Mellstedt,

and A. Osterborg. 2001. Interleukin 4 therapy for patients with chronic lymphocytic leukaemia: a phase I/II study. *Br J Haematol* 112, no. 1:155).

In many systems the effects of IL13 are largely similar to those of IL4, but IL13 is slightly less potent that IL4. It is unclear whether B-CLL cells express IL13, but the cells do express the IL13 receptor. The effects of IL13 in B-CLL are controversial. While Chaouchi et al. suggested that IL13, like IL4 protects B-CLL cells from apoptosis (Chaouchi, N., C. Wallon, C. Goujard, G. Tertian, A. Rudent, D. Caput, P. Ferrera, A. Minty, A. Vazquez, and J.F. Delfraissy. 1996. Interleukin-13 inhibits interleukin-2-induced proliferation and protects chronic lymphocytic leukemia B cells from in vitro apoptosis. *Blood* 87, no. 3:1022), studies by Fluckiger et al. suggest that this is not the case (Fluckiger, A.C., F. Briere, G. Zurawski, J.M. Bridon, and J. Banchereau. 1994. IL13 has only a subset of IL4-like activities on B chronic lymphocytic leukaemia cells. *Immunology* 83, no. 3:397).

The combined finding of 2D and 3D structure similarity to 4-helical cytokines and the importance of IL4 in B-CLL strongly suggests that the novel class of proteins of which the AMB-1 protein is one representative are cytokines.

In one aspect the invention relates to a method of identifying a receptor for an isolated polypeptide as in the present invention, said method comprising the steps of contacting the isolated polypeptide or an expression vector encoding said isolated polypeptide with at least one cell line being dependent on a specific cytokine and observing at least one parameter selected from the group consisting of: proliferation, apoptosis, necrosis, cell cycle changes or other physiological responses. Other parameters: inhibition of /activation of enzymes or caspases, upregulation of/ degradation of mRNA or proteins involved in proliferation, apoptosis, necrosis or cell cycle changes. By knowing the response of the cytokine dependent cell line to known cytokines it is possible to assign a receptor to the polypeptide. This receptor/cytokin match can be confirmed by blocking the receptor with receptorspecific antibodies.

In a further aspect the invention relates to a method of identifying a receptor for an isolated polypeptide as defined in the present invention, said method comprising the steps of contacting the isolated polypeptide with a plurality of polypeptides and

selecting polypeptides that bind to the isolated polypeptide as receptors. This method is more based on the chemical properties of the polypeptides of the present invention.

Still further there is provided a method for identifying a modulator of the binding between an isolated polypeptide according to the present invention and a receptor identified according to any the present invention, said method comprising providing a complex between said polypeptide and said receptor, said complex having a predetermined K_D, and providing a plurality of putative modulators, contacting said complex with said plurality of putative modulators, and selecting those modulators that cause an increase in the KD of at least 10%, more preferably more than 20 %, more preferably more than 50 %, more preferably more than 100 %, more preferably more than 200 %, more preferably more than 5 times, more preferably more than 10 times, such as more than 100 times, for example more than 1000 times, such as more than 10,000 times, for example more than 100,000 times, such as more than 1,000,000 times. These modulators can be used as drug leads in the development of drugs against B-CLL.

In a further aspect there is provided a pharmaceutical composition comprising an isolated polypeptide as defined in the present invention and a pharmaceutically acceptable carrier. The novel class of proteins are expected to have several pharmaceutical uses.

The novel proteins may also be used for the preparation of a medicament for the treatment of bone disorders, inflammation, for lowering blood serum cholesterol, allergy, infection, viral infections, hematopoietic disorders, preneoplastic lesions, immune related diseases, autoimmune related diseases, infectious diseases, tuberculosis, cancer, viral diseases, septic shock, reconstitution of the haematopoietic system, induction of the granulocyte system, pain, cardial dysfunction, CNS disorders, depression, artheritis, psoriasis, dermatitis, collitis, Crohn's disease, and diabetes, in a subject in need thereof.

Further uses of the novel class of proteins include use as a growth factor, use as an adjuvant or as an immune anhancer, use for regulating TH2 immune responses, and use for suppressing Th1 immune responses.

One further therapeutic application of the present invention is a method of vaccination against B-CLL said method comprising immunising a subject against a translational product of SEQ ID No 1. By stimulating the immune system of a subject to produce antibodies against the translational product the subject can become immune towards B-CLL and/or the method can be used as part of therapy. The state of the art describes various ways of immunising a subject against a particular protein.

With the invention of a new class of proteins the invention relates to a method for producing an antibody with specificity against an isolated polypeptide as defined in the present invention, said method comprising the steps of

i) providing a host organism,

ii) immunising said host organism with an isolated polypeptide as defined in the present invention, or transfecting said host organism with an expression vector capable of directing the expression of an isolated polypeptide as defined in the present invention,

iii) obtaining said antibody.

The antibodies obtainable by this method can be used for diagnostic as well as therapeutic applications.

For example the antibodies may be formulates as a pharmaceutical composition comprising an antibody according to the invention and pharmaceutically acceptable carriers. Once the antibodies have been produced in a suitable host cell it is also possible to isolate and/or construct an expression vector encoding said antibody and to use said vector for recombinant production of the antibody. In this way it is possible to produce a human antibody in a high producing cell line such as yeast or bacteria.

In a still further aspect the invention relates to an isolated polynucleotide selected from the group consisting of:

i) a polynucleotide comprising nucleotides 40001 to 60000 of SEQ ID No 1,

ii) a polynucleotide encoding a polypeptide having the amino acid sequence of SEQ ID No 3,

iii) a polynucleotide, the complementary strand of which hybridises, under stringent conditions, with a polynucleotide as defined in any of i) and ii), and encodes a polypeptide, which

a) has at least 60 % sequence identity with the amino acid sequence of SEQ ID No 3 and has interleukin or cytokine activity,

b) is recognised by an antibody, or a binding fragment thereof, which is capable of recognising an epitope, wherein said epitope is comprised within a polypeptide having the amino acid sequence of SEQ ID No 3; and/or

c) is competing with a polypeptide having the amino acid sequence as shown in SEQ ID No 3 for binding to at least one predetermined binding partner such as a cytokine receptor,

iv) a polynucleotide which is degenerate to the polynucleotide of iii), and

v) the complementary strand of any such polynucleotide.

The novelty of the polypeptide sequences according to the present invention arises from the discovery of the present inventors that this polynucleotide encodes a novel class of 4-helical cytokines and the discovery that the expressed parts of such polynucleotides can be used for diagnosis of a subtype of B-CLL. The promoter sequence (which forms part of SEQ ID No 1) and the coding sequences can be used in various aspects of gene therapy and immunotherapy.

Further polynucleotide sequences from other subjects or other species with the same function can be isolated by one of the following methods, which each form independent aspects of the present invention.

A first method for identifying a nucleotide sequence encoding a 4-helical cytokine comprises the steps of:

i) isolating mRNA from a biological sample,

ii) hybridising the mRNA to a probe comprising at least 10 nucleotides of the coding sequence of SEQ ID No 1 (nucleotides no 52051 to 52466) under stringent conditions,

iii) determining the nucleotide sequence of a sequence capable of hybridising under step ii), and

iv) determining the presence of an open reading frame in the nucleotide sequence determined under step iii).

A second method for identifying a nucleotide sequence encoding a 4-helical cytokine is a computer assisted method comprising the steps of
i) performing a sequence similarity search of at least 10 nucleotides of the coding sequence SEQ ID No 1 (nucleotides no 52051 to 52466),
ii) aligning "hits" to said coding sequence,
iii) determining the presence of an open reading frame in the "hits".

It is highly likely that other similar polypeptides encoding further 4-helical cytokines can be found in other subjects and/or other species of mammals. In particular, subjects of other geographical origin may carry genes which differ from the polynucleotides of the present invention. It is also conceivable that similar sequences can be found in closely and even in distantly related species.

In a further aspect of the present invention is provided a method of preparing a 4-helical cytokine, said method comprising the steps of identifying a further polynucleotide sequence ecoding a 4-helical cytokine, and further comprising synthesising the polypeptide encoded by the open reading frame and determining the activity of said polypeptide in a cytokine activity assay, preferably an interleukin assay, more preferably an interleukin-4 assay. Thereby it is ascertained that the isolated polypeptides indeed have 4-helical cytokine activity. In the detailed description and the appended examples there is provided methods for chemical and biological assaying of 4-helical cytokine activity.

The 4-helical cytokines may be used for preparing a pharmaceutical composition by further carrying of the step of formulating the polypeptide with a pharmaceutically acceptable carrier or diluent.

Furthermore there are provided various different methods for screening compounds capable of treating B-CLL. In a first method, screening comprises administering a test-compound to a host cell comprising a recombinant expression construct, said expression construct comprising the promoter sequence of bases no. 40001 to 51417 or 40001 to 49100 of SEQ ID No 1 or a fragment thereof operably linked to a

reporter gene, and determining the presence and/or amount of the reporter gene product. This method is very useful for automated high throughput screening.

In a second screening method, screening comprises administering a test-compound to a host cell comprising a recombinant expression construct, said expression construct comprising a constitutive promoter directing the expression of a polypeptide according to the invention and on said cell measuring a parameter selected from the group consisting of: proliferation, apoptosis, necrosis, cell cycle changes or other physiological responses. Other parameters: inhibition of /activation of enzymes or caspases, upregulation of/ degradation of mRNA or proteins involved in proliferation, apoptosis, necrosis or cell cycle changes.

In a third screening method, screening comprises administering a test-compound to a cell line established from a subject diagnosed according to the invention, said method comprising measuring: proliferation, apoptosis, necrosis, cell cycle changes or other physiological responses. Other parameters: inhibition of /activation of enzymes or caspases, upregulation of/ degradation of mRNA or proteins involved in proliferation, apoptosis, necrosis or cell cycle changes.

Finally, the invention provides a method for determining an increased or decreased predisposition for B-CLL comprising determining in a biological sample from a subject a germline alteration in a target nucleic acid sequence comprising 150,000 nucleotides, said target nucleic acid sequence comprising at least 10 nucleotides of SEQ ID No 1. This aspect is based on the finding of the importance of the expression product of SEQ ID No 1, and the absence of any detectable expression product of SEQ ID No 1 in healthy tissue and in patients with good prognosis B-CLL. It is highly likely that the difference is caused by a germline alteration. A germline alteration can be targeted by gene therapy methods and by the methods provided in the present invention.

Description of Drawings

Figure 1: Overall survival of B-CLL patients by genotype (all stages) The prognostic significance of V_H homology and cytogenetic aberrations is independent of clinical stage (from Kröber et al., 2002 (4)).

Figure 2: RT-PCR performed on 16 B-CLL patients. UPN1-8 are unmutated patients and UPN9-16 are mutated patients.

Figure 3: Northern blot analysis on RNA from blood samples of B-CLL patients and from various tissue and cell line samples. The approximate positions of 18S and 28S rRNA are marked. The probe was an 896 bp fragment obtained by RT-PCR of UPN 7.

Figure 4. Searches with the peptide sequence in the sptrnr data base of peptide sequences (includes Sprot and nrtrembl) showing a similarity to putative intron maturases from cloroplasts and to bovine IL4.

Figure 5. A 3D search, where the peptide sequence has been searched for similarity to known protein or peptide 3D-structures.

Figure 6. Predicted 3-D structure of AMB-1 compared to the known 3-D structure of human IL4. Prediction is performed using SEQ ID No 3 and the method described in: Enhanced Genome Annotation using Structural Profiles in the Program 3D-PSSM. Kelley LA, MacCallum RM & Sternberg MJE (2000). J. Mol. Biol. 299(2), 499-520.

Figure 7. Alignment of the AMB1 peptide sequence with the sequences of IL4, IL3, IL13 and GM-CSF, based on their structures.

Figure 8. Genomic sequence (SEQ ID No 1) of the part of the human chromosome 12 comprising the AMB-1 transcript and the AMB-1 protein. The sequence consists of bases 40,000 to 60,000 of AC063949.emhum. Bold nucleotides correspond to the transcript. The open reading frame of exon 1 encoding the AMB-1 4-helical cytokine (SEQ ID No 3) is shown (nucleotides 52051 to 52466).

Figure 9. AMB1 mRNA Longest form (SEQ ID No 4). Short form (SEQ ID No 2) starts around pos. 2317. Coding region: 3001 – 3363 Stop codon 3364-3366. Position of intron 4254. Intron length 3099 (not included).

Figure 10. The amino acid sequence in one-letter code of the B-CLL associated protein, AMB-1. The sequence is designated SEQ ID No 3.

Figure 11. A table showing the tissue types on the MTE array used for dot blotting of AMB-1 to check for expression in other tissue types.

Detailed description of the invention

Methods of diagnosis

One important aspect of the present invention relates to diagnosis of a subtype of B-cell chronic lymphocytic leukaemia (B-CLL). These methods are based on the discovery by the present inventors that a transcriptional or translational product of SEQ ID No 1 is only present in one particular subtype of B-CLL and completely absent in other subtypes of B-CLL and in healthy tissue (see in particular example 2). By completely absent is meant that the transcriptional or translational products are not detected in any of the other tissue types with the methods used in the appended examples. This is indicative of a complete absence of any transcript or a very low level of transcript in the other tissue types.

The transcriptional product has almost exclusively been found in patients with poor B-CLL prognosis, i.e. in patients with unmutated Ig VH genes. However this finding is based on a limited number of patients so the present inventors expect that it turns out that the subtype of B-CLL is characterised solely or better by the presence of a transcriptional or translational product of SEQ ID No 1. This may in particular be the case when patients from other geographical areas are examined.

Preferably the subject is a mammal, more preferably a human being. It is also expected that the gene encoded by SEQ ID No 1 can be used as a diagnostic tool in other species in particular in mammals selected from the group: domestic animals such as cow, horse, sheep, pig; and pets such as cat or dog.

Preferably, the transcriptional product is a mRNA sequence corresponding to SEQ ID No 2 (short cDNA clone) SEQ ID No 4 (long cDNA clone) or a fragment thereof. Both of these mRNA sequences have been found in patients with poor prognosis.

The mRNA sequence may be detected in a sample using hybridisation techniques. In particular when more than one analysis is to be performed at the same time it is advantageous to use a DNA array comprising an oligomer of at least 20 consecutive bases from the sequence 49101 – 53354 or 56454 – 58408 of SEQ ID No 1.

Another way of detecting the presence or absence of the transcriptional product is by specifically amplifying a transcriptional product having a sequence corresponding to SEQ ID No 2 or 4 or a fragment thereof. This can be done by selecting primer pairs which cause only the amplification of these sequences.

According to another embodiment, the translational product is a protein encoded by SEQ IN No 1 and/or 2 and/or 4. Detection of this protein can be done with state of the art methods including the detection with an antibody directed against said protein, such as Western blotting, more preferably by using a fluorescently labelled antibody, preferably wherein the method comprises the use of FACS. Other methods include gel electrophoresis, gel filtration, ion exchange chromatography, FPLC, Mass spectrometry.

Preferably, said protein is selected from the group comprising SEQ ID No 3 (protein), or a protein sharing at least 60 % sequence identity with SEQ ID No 3. The protein with the amino acid sequence set forth in SEQ ID No 3 is the longest open reading frame in the cDNA sequence of SEQ ID No 2 or 4.

The methods described so-far relate to the determination of the presence or absence of a transcriptional or translational product of SEQ ID No 1. By measuring quantitatively the amount of a transcriptional or translational product of SEQ ID No 1 in a biological sample isolated from a subject, it is possible to predict the progress/stage of B-CLL in a subject.

In one embodiment the quantitative measurement is performed during treatment to estimate the efficiency of such treatment.

For all diagnostic application of the present invention, the biological sample may be selected from the group comprising: a blood sample, lymph node tissue, bone

marrow, or spinal liquid. The cells to be assessed in a sample are leukocytes, mononuclear leukocytes or lymphocytes or B-lymphocytes.

B-CLL therapy

With the identification of a new sub-type of B-CLL the present inventors also provide methods for treatment of B-CLL in such patients. These methods are based on administering to a subject being diagnosed according to the present invention a therapeutically effective amount of a compound capable of selectively killing and/or inhibiting division of and/or inducing apoptosis in B-CLL cells. Preferably the compound is selected from the group chemotherapeutic agents, anti-CD20, anti-CD52- or other antibodies, or the treatment may comprise of non-myeloablative bone marrow transplantation.

In a further therapeutic aspect there is provided a method of treating B-CLL comprising administering to a subject with a B-CLL diagnosis a compound capable of decreasing or inhibiting the formation of a transcriptional and/or translational product from SEQ ID No 1. This method is based on the finding that this transcriptional and/or translational product is only present in B-CLL cells of patients with a poor prognosis and that the protein encoded by SEQ ID No 1 is the etiological factor in B-CLL. By inhibiting the activity of this protein and/or by inhibiting its synthesis a treatment and/or cure for B-CLL is provided.

In one embodiment the compound is a therapeutic antibody directed against a polypeptide having the amino acid sequence of SEQ ID No 3, preferably wherein said antibody is a human or humanised antibody. Another possibility is to identify a modulator of binding of SEQ ID No 3 to its receptor within or outside the cell and to administer this modulator to the cells.

Other methods are aimed at decreasing and/or inhibiting transcription. One method is based on administering an oligonucleotide capable of inhibiting transcription from SEQ ID No 1. Said oligonucleotide may comprises at least 8-10 consecutive nucleotides from the sequence 40001 to 51417 or the sequence 40001 to 49100 of SEQ ID No 1. These sequences constitute the putative promoter sequences of the short and long mRNAs encoding SEQ ID No 3. The oligonucleotides bind

specifically to the promoter sequences and inhibit transcription of the gene. Such oligonucleotides may comprises nucleotide monomers selected from the group: DNA, RNA, LNA, PNA, methylated DNA, methylated RNA, more preferably PNA or LNA.

In a more preferred embodiment the therapeutic methods comprise administering an oligonucleotide capable of binding to a transcriptional product and preventing translation. One particularly preferred embodiment of this aspect is RNAi oligonucleotides. RNAi works by hybridising specifically to the mRNA transcribed by the cell to form a (partly) double stranded RNA molecule. This is recognised as a double stranded molecule by the cell's own nucleases, which degrade them. In order for the technique to work efficiently, the RNAi oligonucleotide comprises 8-22 consecutive nucleotides of the complementary sequence or SEQ ID No 2 and/or SEQ ID No 4, more preferably of SEQ ID No 2. By selecting a sequence from SEQ ID No 2, both mRNAs can be targeted and broken down.

RNAi oligonucleotides may be administered to the cell, or a vector may be transfected into the cells, said vector comprising a promoter region capable of directing the expression of at least one RNAi oligonucleotide. Due to the very restricted expression of the AMB-1 gene, it is not important only to target the RNAi oligos or the vectors to B-CLL cells.

One way of targeting to blood cells comprises using a heparin receptor for targeting to blood cells.

Another way of addressing the transcriptional product of SEQ ID No 1 is to use an antisense construct comprising a promoter sequence capable of directing the transcription of at least part of the antisense equivalent of SEQ ID No 1 or 2 or 4. As for the RNAi oligonucleotides targeting to B-CLL cells is not particularly important.

When desired targeting to B-CLL cells can be performed using the CD19 or CD20 receptor. The CD19 receptor is particularly preferred since it internalises its ligand.

In a further therapeutic embodiment the compound is a gene therapy vector comprising a promoter sequence operably linked to a sequence coding for a protein capable of inhibiting cell division in the cell and/or capable of killing the cell, said promoter sequence being a tissue specific promoter capable of directing expression only in B cells, more preferably only in B-CLL cells. One particularly preferred promoter sequence is the extremely cell specific promoter of SEQ ID No 1. Said promoter sequence comprises bases No 40001 to 51417 of SEQ ID No 1 or a fragment thereof, such as the fragment from 40001 to 49100 or a fragment of this fragment. When this promoter is used targeting of the suicide vector is not very important, since it will only be active in the cells in which AMB-1 is expressed and these are the cells to be targeted by the suicide gene.

Deletion studies will determine the exact length of the promoter sequence counted from the transcription start site. Accordingly, the promoter may comprise at least 100 nucleotides 5' to base no. 51471 or 49100 of SEQ ID No 1, such as at least 200 nucleotides, for example at least 300 nucleotides, such as at least 400 nucleotides, for example at least 500 nucleotides, such as at least 600 nucleotides, for example at least 700 nucleotides, such as at least 800 nucleotides, for example at least 900 nucleotides, such as at least 1000 nucleotides, for example at least 1100 nucleotides, such as at least 1200 nucleotides, for example at least 1300 nucleotides, such as at least 1400 nucleotides, for example at least 1500 nucleotides, such as at least 1600 nucleotides, for example at least 1700 nucleotides, such as at least 1800 nucleotides, for example at least 1900 nucleotides, such as at least 2000 nucleotides, for example at least 2500 nucleotides, such as at least 3000 nucleotides, for example at least 3500 nucleotides, such as at least 5000 nucleotides, for example at least 10,000 nucleotides.

4-helical cytokines

4-helical cytokines of the present invention include isolated polypeptides selected from the group

i) a polypeptide comprising or essentially consisting of the amino acid sequence of SEQ ID No. 3, or a fragment thereof, or

ii) a polypeptide functionally equivalent to SEQ ID No. 3, or a fragment thereof, sharing at least 60 % sequence identity with SEQ ID No 3, wherein said fragment or functionally equivalent polypeptide

a. has interleukin or cytokine activity; and/or

b. is recognised by an antibody, or a binding fragment thereof, which is capable of recognising an epitope, wherein said epitope is comprised within a polypeptide having the amino acid sequence of SEQ ID No 3; and/or

c. is competing with a polypeptide having the amino acid sequence as shown in SEQ ID No 3 for binding to at least one predetermined binding partner.

These polypeptides constitute a novel class of proteins sharing 2D and 3D structure similarities with 4-helical cytokines. In a preferred embodiment, the isolated polypeptide comprises or essentially consists of the amino acid sequence of SEQ ID No. 3 or a fragment thereof. This is the protein found to be expressed solely in B-CLL cells of patients having a poor prognosis. This particular protein at least can be used for diagnosis, for raising antibodies for use in therapy against B-CLL, and for protective or therapeutic immunisation of a subject against B-CLL.

The protein defined by SEQ ID No 3 shares very little sequence identity with known cytokines and interleukines and as a matter of fact very little sequence identity with any known protein. Consequently the present inventors contemplates that the group comprises functionally equivalent polypeptide sharing at least 60% sequence identity with SEQ ID No 3, more preferably at least 70% sequence identity, more preferably at least 80 % sequence identity, such as at least 90 % sequence identity, for example at least 95 % sequence identity, such as at least 97 % sequence identity, for example at least 98 % sequence identity.

It is expected that the isolated polypeptide have cytokine and/or interleukin activity. Therefore the binding partner of item c) is preferably selected from the group: an antibody directed against SEQ ID No 3, the receptor for IL4, IL3, IL13, GM-CSF, TGF-β, or IGF. Activity as a cytokine or interleukin can also be assessed in a biological assay where the polypeptide is contacted with a cytokine dependent cell line.

Consequently, the isolated polypeptide preferably has interleukin activity, such as having IL3, IL13, GM-CSF, TGF-β, IGF activity, more preferably having IL4 activity.

Probably the isolated polypeptides are capable of forming homo- or hetero-oligomer with each other and among themselves. Such oligomers are also within the scope of the present invention. Such oligomers may comprise at least one isolated polypeptides as defined in any the present invention, such as a dimer, a trimer, a quatramer, a quintamer, a hexamer, an octamer, a decamer, a dodecamer. In biological systems the activity may be attributed only to dimer or higher –mer.

Functional Equivalents

Modification and changes may be made in the structure of the peptides of the present invention and DNA segments which encode them and still obtain a functional molecule that encodes a protein or peptide with desirable characteristics. The following is a discussion based upon changing the amino acids of a protein to create an equivalent, or even an improved, second-generation molecule. The amino acid changes may be achieved by changing the codons of the DNA sequence, according to the genetic code.

For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies, binding sites of receptors, or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence, and, of course, its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. It is thus contemplated by the inventors that various changes may be made in the peptide sequences of the disclosed compositions, or corresponding DNA sequences which encode said peptides without appreciable loss of their biological utility or activity.

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biologic

function on a protein is generally understood in the art (Kyte and Doolittle, 1982, incorporate herein by reference). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like. Each amino acid has been assigned a hydropathic index on the basis of their hydrophobicity and charge characteristics (Kyte and Doolittle, 1982), these are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydropathic index or score and still result in a protein with similar biological activity, ie. still obtain a biological functionally equivalent protein. In making such changes, the substitution of amino acids whose hydropathic indices are within ±2 is preferred, those which are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred. It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U.S. Pat. No. 4,554,101, incorporated herein by reference, states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein.

As detailed in U.S. Pat. No. 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0±1); glutamate (+3.0±1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5±1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4). It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within

±2 is preferred, those which are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions which take various of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

Functional equivalents and variants are used interchangably herein. In one preferred embodiment of the invention there is also provided variants of a 4-helical cytokine, and variants of fragments thereof. When being polypeptides, variants are determined on the basis of their degree of identity or their homology with a predetermined amino acid sequence, said predetermined amino acid sequence being SEQ ID No. 3 or a fragment thereof.

Accordingly, variants preferably have at least 60 % sequence identity, for example at least 65% sequence identity, such as at least 70 % sequence identity, for example at least 75% sequence identity, for example at least 80% sequence identity, such as at least 85 % sequence identity, for example at least 90 % sequence identity, such as at least 91 % sequence identity, for example at least 91% sequence identity, such as at least 92 % sequence identity, for example at least 93 % sequence identity, such as at least 94 % sequence identity, for example at least 95 % sequence identity, such as at least 96 % sequence identity, for example at least 97% sequence identity, such as at least 98 % sequence identity, for example 99% sequence identity with the predetermined sequence.

A degree of identity of amino acid sequences is a function of the number of identical amino acids at positions shared by the amino acid sequences. A degree of homology or similarity of amino acid sequences is a function of the number of amino acids, i.e. structurally related, at positions shared by the amino acid sequences. Sequence identity is determined in one embodiment by utilising fragments of 4-helical cytokines comprising at least 25 contiguous amino acids and having an amino acid sequence which is at least 80%, such as 85%, for example 90%, such

as 95%, for example 99% identical to the amino acid sequence of SEQ ID No. 3, wherein the percent identity is determined with the algorithm GAP, BESTFIT, or FASTA in the Wisconsin Genetics Software Package Release 7.0, using default gap weights.

An "unrelated" or "non-homologous" sequence shares less than 40% identity, though preferably less than 25% identity, with one of the 4-helical cytokine sequences of the present invention. The term "substantial identity" means that two peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 80 percent sequence identity, preferably at least 90 percent sequence identity, more preferably at least 95 percent sequence identity or more (e.g., 99 percent sequence identity). Preferably, residue positions which are not identical differ by conservative amino acid substitutions.

Additionally, variants are also determined based on a predetermined number of conservative amino acid substitutions as defined herein below. Conservative amino acid substitution as used herein relates to the substitution of one amino acid (within a predetermined group of amino acids) for another amino acid (within the same group), wherein the amino acids exhibit similar or substantially similar characteristics.

Within the meaning of the term "conservative amino acid substitution" as applied herein, one amino acid may be substituted for another within the groups of amino acids indicated herein below:

i) Amino acids having polar side chains (Asp, Glu, Lys, Arg, His, Asn, Gln, Ser, Thr, Tyr, and Cys,)

ii) Amino acids having non-polar side chains (Gly, Ala, Val, Leu, Ile, Phe, Trp, Pro, and Met)

iii) Amino acids having aliphatic side chains (Gly, Ala Val, Leu, Ile)

iv) Amino acids having cyclic side chains (Phe, Tyr, Trp, His, Pro)

v) Amino acids having aromatic side chains (Phe, Tyr, Trp)

vi) Amino acids having acidic side chains (Asp, Glu)

vii) Amino acids having basic side chains (Lys, Arg, His)

viii) Amino acids having amide side chains (Asn, Gln)

ix) Amino acids having hydroxy side chains (Ser, Thr)

x) Amino acids having sulphor-containing side chains (Cys, Met),

xi) Neutral, weakly hydrophobic amino acids (Pro, Ala, Gly, Ser, Thr)

xii) Hydrophilic, acidic amino acids (Gln, Asn, Glu, Asp), and

xiii) Hydrophobic amino acids (Leu, Ile, Val)

Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

Accordingly, a variant or a fragment thereof according to the invention may comprise, within the same variant of the sequence or fragments thereof, or among different variants of the sequence or fragments thereof, at least one substitution, such as a plurality of substitutions introduced independently of one another.

It is clear from the above outline that the same variant or fragment thereof may comprise more than one conservative amino acid substitution from more than one group of conservative amino acids as defined herein above.

The addition or deletion of at least one amino acid may be an addition or deletion of from preferably 2 to 250 amino acids, such as from 10 to 20 amino acids, for example from 20 to 30 amino acids, such as from 40 to 50 amino acids. However, additions or deletions of more than 50 amino acids, such as additions from 50 to 100

amino acids, addition of 100 to 150 amino acids, addition of 150-250 amino acids, are also comprised within the present invention. The deletion and/or the addition may - independently of one another - be a deletion and/or an addition within a sequence and/or at the end of a sequence.

The polypeptide fragments according to the present invention, including any functional equivalents thereof, may in one embodiment comprise less than 250 amino acid residues, such as less than 240 amino acid residues, for example less than 225 amino acid residues, such as less than 200 amino acid residues, for example less than 180 amino acid residues, such as less than 160 amino acid residues, for example less than 150 amino acid residues, such as less than 140 amino acid residues, for example less than 130 amino acid residues, such as less than 120 amino acid residues, for example less than 110 amino acid residues, such as less than 100 amino acid residues, for example less than 90 amino acid residues, such as less than 85 amino acid residues, for example less than 80 amino acid residues, such as less than 75 amino acid residues, for example less than 70 amino acid residues, such as less than 65 amino acid residues, for example less than 60 amino acid residues, such as less than 55 amino acid residues, for example less than 50 amino acid residues.

"Functional equivalency" as used in the present invention is according to one preferred embodiment established by means of reference to the corresponding functionality of a predetermined fragment of the sequence.

Functional equivalents or variants of a 4-helical cytokine will be understood to exhibit amino acid sequences gradually differing from the preferred predetermined 4-helical cytokine, as the number and scope of insertions, deletions and substitutions including conservative substitutions increases. This difference is measured as a reduction in homology between the preferred predetermined sequence and the fragment or functional equivalent.

All fragments or functional equivalents of SEQ ID No. 3 are included within the scope of this invention, regardless of the degree of homology that they show to the respective, predetermined 4-helical cytokines disclosed herein. The reason for this is that some regions of the 4-helical cytokines are most likely readily mutatable, or

capable of being completely deleted, without any significant effect on the binding activity of the resulting fragment.

A functional variant obtained by substitution may well exhibit some form or degree of native cytokine activity, and yet be less homologous, if residues containing functionally similar amino acid side chains are substituted. Functionally similar in this respect refers to dominant characteristics of the side chains such as hydrophobic, basic, neutral or acidic, or the presence or absence of steric bulk. Accordingly, in one embodiment of the invention, the degree of identity is not a principal measure of a fragment being a variant or functional equivalent of a preferred predetermined fragment according to the present invention.

One particularly preferred method of determining the degree of functional equivalence is by performing a biological or chemical assay such as the assays described in the appended examples. Preferred functional equivalents of SEQ ID No 3 are those that have a K_D with respect to a predefined receptor which is less than 10 times higher than the K_D of the polypeptide of SEQ ID No 1 with respect to the same receptor, more preferably less than 5 times higher, more preferably less than 2 times higher.

With respect to functional equivalence this may be defined in a biological assay based on a cytokine dependent or stimulated cell line. Such cell lines are e.g. available from American Type Culture Collection, P.O.Box 1549, Manassas, VA 20108 USA. The following cell lines at least are available for testing cytokines and in particular interleukins:

Accession number	**Description**	**Activity**
CRL-1841	TH-2 clone A5E	IL2 dependent, IL4 stimulated
CRL-2003	TF-1	IL3 dependent
CRL-2407	NK-92	IL2 dependent
CRL-2408	NK-92MI	IL2 dependent
CRL-2409	NK92CI	IL2 dependent
CRL-9589	AML-193	IL3 stimulated, GM-CSF dep.
CRL-9591	MV-4-11	GM-CSF dependent
TIB-214	CTLL-2	IL2 dependent

TIB-239 2E8 IL7 dependent

The following cell lines are available from DSMZ - Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, D-38124 Braunschweig, GERMANY. As can be seen from the table, some of the cell lines can be used to broadly assess cytokine activity whereas others are only reported to respond to one or a few specific cytokines.

Accession number	Description	Acvitity
ACC 211	Mouse hybridoma, B9	IL6 dependent
ACC 137	Human acute myeloid leukemia, UT-7	Constitutively cytokine responsive to various cytokines.
ACC 104	Human acute megakaryoblastic leukemia	Respond with proliferation to: GM-CSF, IFN-alpha, IFN-á, IFN-gamma, IL2, IL3, IL4, IL6, IL15, NG F, SCF, TNF-alpha, TPO
ACC 247	Human acute myeloid leukemia, OCI-AML5	G-GSF, GM-CSF, IL3, FTL3-ligand
ACC 271	Human acute myeloid leukemia, MUTZ-2	IL3, SCF, G-CSF, M-CSF, IFN-gamma
ACC 334	Human erythroleukemia, TF-1	GM-CSF, IFN-gamma, IL3, IL4, IL5, IL6, IL13, LIF, NGF, OSM, SCF, TNF-alpha, and TPO

The TF-1 cell line mentioned above can be used for assaying IL13 function. This cell line is sensitive to various different cytokines but gives a very strong proliferative response when exposed to IL13. This cell line can in particular be used if there is no response in the IL4 sensitive cell line (CT.h4S). Further cell lines which can be used for distinguishing between IL4 and IL13 activity include cell lines/hybridomas such as B-9-1-3 (Bouteiller, C.L., R. Astruc, A. Minty, P. Ferrara, and J.H. Lupker. 1995. Isolation of an IL13-dependent subclone of the B9 cell line useful for the estimation

of human IL13 bioactivity. *J Immunol Methods* 181, no. 1:29) and A201.1 (Andrews, R., L. Rosa, M. Daines, and G. Khurana Hershey. 2001. Reconstitution of a functional human type II IL4/IL13 receptor in mouse B cells: demonstration of species specificity. *J Immunol* 166, no. 3:1716)

Other chemical modifications of 4-helical cytokines

In addition to the peptidyl compounds described herein, sterically similar compounds may be formulated to mimic the key portions of the peptide structure and that such compounds may also be used in the same manner as the peptides of the invention. This may be achieved by techniques of modelling and chemical designing known to those of skill in the art. For example, esterification and other alkylations may be employed to modify the amino terminus of, e.g., a di-arginine peptide backbone, to mimic a tetra peptide structure. It will be understood that all such sterically similar constructs fall within the scope of the present invention.

Peptides with N-terminal alkylations and C-terminal esterifications are also encompassed within the present invention. Functional equivalents also comprise glycosylated and covalent or aggregative conjugates formed with the same or other 4-helical cytokine fragments and/or 4-helical cytokine molecules, including dimers or unrelated chemical moieties. Such functional equivalents are prepared by linkage of functionalities to groups which are found in fragment including at any one or both of the N- and C-termini, by means known in the art.

Functional equivalents may thus comprise fragments conjugated to aliphatic or acyl esters or amides of the carboxyl terminus, alkylamines or residues containing carboxyl side chains, e.g., conjugates to alkylamines at aspartic acid residues; O-acyl derivatives of hydroxyl group-containing residues and N-acyl derivatives of the amino terminal amino acid or amino-group containing residues, e.g. conjugates with fMet-Leu-Phe or immunogenic proteins. Derivatives of the acyl groups are selected from the group of alkyl-moieties (including C3 to C10 normal alkyl), thereby forming alkanoyl species, and carbocyclic or heterocyclic compounds, thereby forming aroyl species. The reactive groups preferably are difunctional compounds known per se for use in cross-linking proteins to insoluble matrices through reactive side groups.

Covalent or aggregative functional equivalents and derivatives thereof are useful as reagents in immunoassays or for affinity purification procedures. For example, a fragment of 4-helical cytokine according to the present invention may be insolubilised by covalent bonding to cyanogen bromide-activated Sepharose by methods known per se or adsorbed to polyolefin surfaces, either with or without glutaraldehyde cross-linking, for use in an assay or purification of anti-4-helical cytokine antibodies or cell surface receptors. Fragments may also be labelled with a detectable group, e.g., radioiodinated by the chloramine T procedure, covalently bound to rare earth chelates or conjugated to another fluorescent moiety for use in e.g. diagnostic assays.

Synthesis of a 4-helical cytokine

In one embodiment the fragment of 4-helical cytokine is synthesised by automated synthesis. Any of the commercially available solid-phase techniques may be employed, such as the Merrifield solid phase synthesis method, in which amino acids are sequentially added to a growing amino acid chain. (See Merrifield, J. Am. Chem. Soc. 85:2149-2146, 1963).

Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Applied Biosystems, Inc. of Foster City, Calif., and may generally be operated according to the manufacturer's instructions. Solid phase synthesis will enable the incorporation of desirable amino acid substitutions into any fragment of 4-helical cytokine according to the present invention. It will be understood that substitutions, deletions, insertions or any subcombination thereof may be combined to arrive at a final sequence of a functional equivalent. Insertions shall be understood to include amino-terminal and/or carboxyl-terminal fusions, e.g. with a hydrophobic or immunogenic protein or a carrier such as any polypeptide or scaffold structure capable as serving as a carrier.

Oligomers including dimers including homodimers and heterodimers of fragments of 4-helical cytokine according to the invention are also provided and fall under the scope of the invention. 4-helical cytokine functional equivalents and variants can be produced as homodimers or heterodimers with other amino acid sequences or with native 4-helical cytokine sequences. Heterodimers include dimers containing

immunoreactive 4-helical cytokine fragments as well as 4-helical cytokine fragments that need not have or exert any biological activity.

4-helical cytokine fragments according to the invention may be synthesised both in vitro and in vivo. Method for in vitro synthesis are well known, and methods being suitable or suitably adaptable to the synthesis in vivo of 4-helical cytokine are also described in the prior art. When synthesized in vivo, a host cell is transformed with vectors containing DNA encoding 4-helical cytokine or a fragment thereof. A vector is defined as a replicable nucleic acid construct. Vectors are used to mediate expression of 4-helical cytokine. An expression vector is a replicable DNA construct in which a nucleic acid sequence encoding the predetermined 4-helical cytokine fragment, or any functional equivalent thereof that can be expressed in vivo, is operably linked to suitable control sequences capable of effecting the expression of the fragment or equivalent in a suitable host. Such control sequences are well known in the art.

Cultures of cells derived from any organism, prokaryot or eukaryot can be used for expressing the polypeptide. Preferred species are those for which in-vitro protocols are available. Among the bacteria this is particularly the case for E. coli. In principle, any higher eukaryotic cell culture is workable, whether from vertebrate or invertebrate culture. Examples of useful host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cell lines, and WI38, BHK, COS-7, 293 and MDCK cell lines. Preferred host cells are eukaryotic cells known to synthesize endogenous 4-helical cytokine. Cultures of any host cells may be isolated and used as a source of the fragment, or used in therapeutic methods of treatment, including therapeutic methods aimed at promoting or inhibiting a growth state, or screening methods aimed.

Pharmaceutical uses of isolated polypeptides

Apart from being used for diagnosis, it is also within the scope of the present invention to use an isolated polypeptide as defined in the invention for a pharmaceutical composition together with a pharmaceutically acceptable carrier. Such pharmaceutical compositions may be used for any of the purposes for which cytokines and in particular interleukin is used at present.

Examples of such uses include the treatment of bone disorders, inflammation, for lowering blood serum cholesterol, allergy, infection, viral infections, hematopoietic disorders, preneoplastic lesions, immune related diseases, autoimmune related diseases, infectious diseases, tuberculosis, cancer, viral diseases, septic shock, reconstitution of the haematopoietic system, induction of the granulocyte system, pain, cardial dysfunction, CNS disorders, depression, artheritis, psoriasis, dermatitis, collitis, Chron's disease, diabetes, in a subject in need thereof.

It is also within the scope of the present invention to use an isolated polypeptide according to the invention as an adjuvant or as an immune anhancer, for regulating TH2 immune responses, and for suppressing Th1 immune responses.

A further use of an isolated polypeptide of the invention is as a growth factor for administration to cell cultures or as a growth factor for veterinary use, e.g. for stimulating the growth of livestock.

Immunotherapy

Having identified a transcriptional and/or translational product of SEQ ID No 1 as an etiological factor in B-CLL it is also within the scope of the present invention to perform vaccination against B-CLL by immunising a subject against a translational product of SEQ ID No 1. In this way the subject builds up antibodies directed against said translational product and any developing B-CLL will be stopped by these antibodies.

Immunisation may be performed in various ways. such as by immunising said subject with at least one isolated polypeptide as defined the present invention and optionally adjuvants and carriers or immunising with an expression construct capable of expressing an isolated polypeptide according to the invention in the cells (DNA vaccination).

Another method comprises peptide loading of dendritic cells, or ex vivo expansion and activation of T-cells, or inducing a CTL response that targets cells expressing the polypeptide encoded by SEQ ID No 1.

Antibodies

Antibodies against any of the polypeptides belonging to the novel class of proteins identified by the present inventors can be produced by any known method of immunisation.

In one embodiment, the antibodies are produced in a non-human mammal, or in an insect. If antibodies are to be used for therapy in human beings they are preferably subsequently humanised. In one embodiment, the antibody is formulated into a single-chain antibody.

In another embodiment, in particular for therapeutic purposes, the host organism is a human being and the antibody is subsequently produced recombinantly in a non-human mammal, such as a mouse. The antibody may also be produced as a monoclonal antibody in a hybridoma.

The antibodies of the present invention may be provides as part of a pharmaceutical composition. Such a pharmaceutical composition may be used for treating cancer, preferably for treating leukaemia, more preferably for treating B-CLL leukaemia, more preferably for treating poor prognosis B-CLL leukaemia.

Antibodies: Definitions

Adjuvant: Any substance whose admixture with an administered immunogenic determinant increases or otherwise modifies the immune response to said determinant.

Antibody: Immunoglobulin molecule or immunologically active portion thereof, i.e. molecules that contain an "antigen binding site" or paratope. An antigen binding site is that structural portion of an antibody molecule that specifically binds to an antigen at a B cell epitope.

Antibody fragment refers to a portion of a full-length antibody, generally the antigen binding or variable region. Examples of antibody fragments include Fab, Fab', F(ab')

$_2$ and Fv fragments. Papain digestion of antibodies produces two identical antigen binding fragments, called the Fab fragment, each with a single antigen binding site, and a residual "Fc" fragment, so-called for its ability to crystallize readily. Pepsin treatment yields an F(ab') $_2$ fragment that has two antigen binding fragments which are capable of cross-linking antigen, and a residual other fragment (which is termed pFc'). Additional fragments can include diabodies, linear antibodies, single-chain antibody molecules, and multispecific antibodies formed from antibody fragments. As used herein, "functional fragment" with respect to antibodies, refers to Fv, F(ab) and F(ab')$_2$ fragments.

Antibody fragments retain some ability to selectively binding with its antigen or receptor and are defined as follows:

Fab is the fragment that contains a monovalent antigen-binding fragment of an antibody molecule. A Fab fragment can be produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain.

Fab' is the fragment of an antibody molecule and can be obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion of the heavy chain. Two Fab' fragments are obtained per antibody molecule. Fab' fragments differ from Fab fragments by the addition of a few residues at the carboxyl terminus of the heavy chain CH1 domain including one or more cysteines from the antibody hinge region.

(Fab')$_2$ is the fragment of an antibody that can be obtained by treating whole antibody with the enzyme pepsin without subsequent reduction. F(ab')$_2$ is a dimer of two Fab' fragments held together by two disulfide bonds.

Fv is the minimum antibody fragment that contains a complete antigen recognition and binding site. This region consists of a dimer of one heavy and one light chain variable domain in a tight, non-covalent association (V_H -V_L dimer). It is in this configuration that the three CDRs of each variable domain interact to define an antigen binding site on the surface of the V_H -V_L dimer. Collectively, the six CDRs confer antigen binding specificity to the antibody. However, even a single variable

domain (or half of an Fv comprising only three CDRs specific for an antigen) has the ability to recognize and bind antigen, although at a lower affinity than the entire binding site.

Single chain antibody ("SCA"), defined as a genetically engineered molecule containing the variable region of the light chain, the variable region of the heavy chain, linked by a suitable polypeptide linker as a genetically fused single chain molecule. Such single chain antibodies are also refered to as "single-chain Fv" or "sFv" antibody fragments. Generally, the Fv polypeptide further comprises a polypeptide linker between the VH and VL domains that enables the sFv to form the desired structure for antigen binding.

Antibody response: Response at least involving the binding of molecularly distinct Ig molecules to different epitopes present on at least one antigen.

Antigenic: Functionality associated with a molecule capable of eliciting an antibody response.

Antigenic determinant: A molecule, or a part thereof, containing one or more epitopes that will elicit an antibody response in a host organism.

Carrier protein: A scaffold structure, e.g. a polypeptide or a polysaccharide, to which an immunogenic determinant is capable of being associated.

Complement: A complex series of blood proteins whose action "complements" the work of antibodies. Complement destroys bacteria, produces inflammation, and regulates immune reactions.

Conjugated: An association formed between an immunogenic determinant and a carrier. The association may be a physical association generated e.g. by the formation of a chemical bond, such as e.g. a covalent bond, formed between the immunogenic determinant and the carrier.

Co-immunisation: Immunisation by means of separate and/or sequential administration to an individual of an immunogenic determinant and a carrier.

Cytokine: Growth or differentiation modulator, used non-determinative herein, and should not limit the interpretation of the present invention and claims. In addition to the cytokines, adhesion or accessory molecules, or any combination thereof, may be employed alone or in combination with the cytokines.

Cytotoxic response: T-cell mediated destruction of a target cell.

Effective amount: An effective amount of an immunostimulating fragment of TGF-□ sufficient to enhance a humoral and/or cellular immune response induced by an immunogenic composition including a vaccine.

Enhancing immunity is in reference to an animal's response to an antigen expressed by a cell refers to an increase in the level of the animal's immune response to the antigen. The level of an animal's immune response may be measured by, for example, isolating MHC class I-restricted cytotoxic T lymphocytes (CTL) from an animal harboring cells which express the antigen, contacting these CTL cells in vitro with cells expressing the antigen, and determining the cytolytic activity of the CTL cells. Alternatively, where the antigen is expressed by a tumor cell, the level of an animal's immune response to the antigen may be determined in vivo by measuring tumor incidence, the time period between administration of antigen-expressing tumor cells and the development of tumors, and rate of increase in tumor size (e.g., tumor diameter or volume).

Epitope: A specific site on a protein to which only certain antibodies bind.

Hapten: A compound, usually of low molecular weight, that is not in itself immunogenic but that, after administration with a carrier protein or cells (either conjugated or non-conjugated), becomes immunogenic and induces an antibody response resulting in antibody binding of the hapten in the absence of carrier.

Immunization: Process of inducing an immunological response in an organism.

Immunogenic determinant: A molecule, or a part thereof, containing one or more epitopes that will stimulate the immune system of a host organism to make a

secretory, humoral and/or cellular antigen-specific response, or to a DNA molecule which is capable of producing such an immunogen in a vertebrate.

Immunological response: Response to a immunogenic composition comprising an immunogenic determinant. An immune response involves the development in the host of a cellular- and/or antibody-mediated response to the administered composition or vaccine in question. An immune response generally involves the action of one or more of i) the antibodies raised, ii) B cells, iii) helper T cells, iv) suppressor T cells, and v) cytotoxic T cells, directed specifically to an immunogenic determinant present in an administered immunogenic composition.

Immunogenic composition: Composition capable of raising an immunological response in an individual.

Immunogenic: Functionality associated with an entity capable of eliciting an immunological response.

Immunostimulating effect: Functionality associated with an entity capable of eliciting an enhanced immune response. An enhanced immune response will be understood within the meaning of the observed difference in the immune response measured as an enhancement of an antibody production and/or a cytotoxic T-cell activity, or otherwise registered, when an immunogenic composition is administered in the presence or absence, respectively, of the entity. An immunogenic composition comprising the entity will be understood as being a composition according to the present invention.

Increased level of presentation of an antigen on a cell surface by an MHC class I molecule refers to a quantity of the antigen which is physically associated (e.g., non-covalently) with a cell surface-bound MHC class I molecule and which is greater than a quantity of the antigen associated with the cell surface-bound MHC class I molecule in a corresponding control cell. An increase in the level of presentation of an antigen in a cell refers to a quantity of the antigen which is physically associated with a cell surface-bound MHC class I molecule which is greater than the quantity of the antigen which is physically associated with the cell surface-bound MHC class I molecule in a corresponding control cell, preferably about two-fold greater than,

more preferably about three-fold greater than, and most preferably at least about five-fold greater than the quantity of the MHC class I molecule in a corresponding control cell. The level of presentation of an antigen by an MHC class I molecule may readily be determined by, for example, flow cytometric analysis as described herein.

MHC (major histocompatible complex): The term "MHC class I molecule" refers to a glycoprotein which is integral to the cell membrane. An MHC class I molecule is composed of two polypeptide chain, i.e., a transmembrane polypeptide of approximately Mr 45K which is noncovalently associated with a nonpolymorphic extracellular polypeptide, β_2 -microglobulin. The transmembrane polypeptide is composed of an extracellular domain, a hydrophobic transmembrane domain and a cytoplasmic domain. One of the most important functions of MHC class I molecule is to present, on the cell surface, antigenic peptide fragments of intracellularly generated foreign protein antigens in a form that T cells can recognize. For example, an MHC class I molecule forms a complex with a viral antigen which is processed and degraded intracellularly to a short peptide fragment, and the formed complex is recognized as 'altered self' MHC and bound by a T cell receptor on a cytotoxic T cell as the first step in triggering lysis of a virus-infected cell. Similarly, as part of tumor surveillance, tumor-associated antigens also bind to MHC class I molecules on the membrane surface of neoplastic cells to form a complex which is recognized by cytotoxic lymphocytes, resulting in lysis of the neoplastic cell. Examples of MHC class I molecules include murine H-2K and H-2D, and human HLA-A, HLA-B and HLA-C.

Monoclonal antibody is an antibody produced by a hybridoma cell. Methods of making monoclonal antibody-synthesizing hybridoma cells are well known to those skilled in the art, e.g, by the fusion of an antibody producing B lymphocyte with an immortalized B-lymphocyte cell line.

Polyclonal antibody is a mixture of antibody molecules (specific for a given antigen) that has been purified from an immunized (to that given antigen) animal's blood. Such antibodies are polyclonal in that they are the products of many different populations of antibody-producing cells.

Vaccination: Process of inducing a protective immune response in an organism.

Vaccine: Immunogenic composition capable of raising a protective immune response in a subject.

Use of antibodies in therapy

Antibodies directed against epitopes can be used for prevention and/or therapy of for example B-CLL. Antigenic epitopes can be used as vaccines to stimulate an immunological response in a mammal that is directed against cells having the B-CLL-associated epitope found in AMB-1 protein or functional equivalents. Antibodies directed against the antigenic epitopes of the invention can combat or prevent B-CLL.

An antigenic epitope may be administered to the mammal in an amount sufficient to stimulate an immunological response against the antigenic epitope. The antigenic epitope may be combined in a therapeutic composition and administered in several doses over a period of time that optimizes the immunological response of the mammal. Such an immunological response can be detected and monitored by observing whether antibodies directed against the epitopes of the invention are present in the bloodstream of the mammal.

Such antibodies can be used alone or coupled to, or combined with, therapeutically useful agents. Antibodies can be administered to mammals suffering from any B-CLL that displays the B-CLL-associated epitope. Such administration can provide both therapeutic treatment, and prophylactic or preventative measures. For example, therapeutic methods can be used to determine the spread of a B-CLL and lead to its remission.

Therapeutically useful agents include, for example, leukeran, adrimycin, aminoglutethimide, aminopterin, azathioprine, bleomycin sulfate, bulsulfan, carboplatin, carminomycin, carmustine, chlorambucil, cisplatin, cyclophosphamide, cyclosporine, cytarabidine, cytosine arabinoside, cytoxin dacarbazine, dactinomycin, daunomycin, daunorubicin, doxorubicin, esperamicins, etoposide, fluorouracil, ifosfamide, interferon-α, lomustine, melphalan, mercaptopurine, methotrexate, mitomycin C, mitotane, mitoxantrone, procarbazine HCl, taxol, taxotere (docetaxel),

teniposide, thioguanine, thiotepa, vinblastine sulfate, vincristine sulfate and vinorelbine. Additional agents include those disclosed in Chapter 52, Antineoplastic Agents (Paul Calabresi and Bruce A. Chabner), and the introduction thereto, pp.1202-1263, of Goodman and Gilman's "The Pharmacological Basis of Therapeutics", Eighth Edition, 1990, McGraw-Hill, Inc. (Health Professions Division). Toxins can be proteins such as, for example, pokeweed anti-viral protein, cholera toxin, pertussis toxin, ricin, gelonin, abrin, diphtheria exotoxin, or Pseudomonas exotoxin. Toxin moieties can also be high energy-emitting radionuclides such as cobalt-60, I-131, I-125, Y-90 and Re-186, and enzymatically active toxins of bacterial, fungal, plant or animal origin, or fragments thereof.

Chemotherapeutic agents can be used to reduce the growth or spread of B-CLL cells and tumors that express the AMB-1 associated epitope of the invention. Animals that can be treated by the chemotherapeutic agents of the invention include humans, non-human primates, cows, horses, pigs, sheep, goats, dogs, cats, rodents and the like. In all embodiments human B-CLL antigens and human subjects are preferred.

Species-dependent antibodies can be used in therapeutic methods. Such a species-dependent antibody has constant regions that are substantially non-immunologically reactive with the chosen species. Such species-dependent antibody is particularly useful for therapy because it gives rise to substantially no immunological reactions. The species-dependent antibody can be of any of the various types of antibodies as defined above, but preferably is mammalian, and more preferably is a humanized or human antibody.

Compositions

Therapeutically useful agents can be formulated into a composition with the antibodies of the invention and need not be directly attached to the antibodies of the invention. However, in some embodiments, therapeutically useful agents are attached to the antibodies of the invention using methods available to one of skill in the art, for example, standard coupling procedures.

Compositions may contain antibodies, antigenic epitopes or trypsin-like protease inhibitors. Such compositions are useful for detecting the AMB-1 protein (for example antigenic epitopes) and for therapeutic methods involving prevention and treatment of B-CLLs associated with the presence of the AMB-1 (for example antigenic epitopes).

The antibodies, (and for example antigenic epitopes and protease inhibitors) can be formulated as pharmaceutical compositions and administered to a mammalian host, such as a human patient in a variety of forms adapted to the chosen route of administration. Routes for administration include, for example, intravenous, intra-arterial, subcutaneous, intramuscular, intraperitoneal and other routes selected by one of skill in the art.

Solutions of the antibodies, (and for example antigenic epitopes and protease inhibitors) can be prepared in water or saline, and optionally mixed with a nontoxic surfactant. Formulations for intravenous or intra-arterial administration may include sterile aqueous solutions that may also contain buffers, liposomes, diluents and other suitable additives.

The pharmaceutical dosage forms suitable for injection or infusion can include sterile aqueous solutions or dispersions comprising the active ingredient that are adapted for administration by encapsulation in liposomes. In all cases, the ultimate dosage form must be sterile, fluid and stable under the conditions of manufacture and storage.

Sterile injectable solutions are prepared by incorporating the antibodies, antigenic epitopes and protease inhibitors in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filter sterilization.

Polynucleotides

The isolated polynucleotide of the present invention include the group consisting of:

i. a polynucleotide comprising nucleotides 40001 to 60000 of SEQ ID No 1,

ii. a polynucleotide encoding a polypeptide having the amino acid sequence of SEQ ID No 3,

iii. a polynucleotide, the complementary strand of which hybridises, under stringent conditions, with a polynucleotide as defined in any of i) and ii), and encodes a polypeptide, which

a) has at least 60 % sequence identity with the amino acid sequence of SEQ ID No 3 and has interleukin or cytokine activity,

b) is recognised by an antibody, or a binding fragment thereof, which is capable of recognising an epitope, wherein said epitope is comprised within a polypeptide having the amino acid sequence of SEQ ID No 3; and/or

c) is competing with a polypeptide having the amino acid sequence as shown in SEQ ID No 3 for binding to at least one predetermined binding partner such as a cytokine receptor,

iv. a polynucleotide which is degenerate to the polynucleotide of iii), and

v. the complementary strand of any such polynucleotide.

Specific examples of fragments of SEQ ID No 1 include the nucleotide sequence of SEQ ID No 2 and the nucleotide sequence of SEQ ID No 4.

Further nucleotide sequences encoding 4-helical cytokines may be obtained by in vitro screening or by in silico screening.

In vitro screening comprises comprising the steps of:

i. isolating mRNA from a biological sample,

ii. hybridising the mRNA to a probe comprising at least 10 nucleotides of the coding sequence of SEQ ID No 1 (nucleotides no 52051 to 52466) under stringent conditions,

iii. determining the nucleotide sequence of a sequence capable of hybridising under step ii), and

iv. determining the presence of an open reading frame in the nucleotide sequence determined under step iii).

Preferably the the open reading frame encodes a polypeptide having at least 60 % sequence identity with the amino acid sequence of SEQ ID No 3. More preferably the sequence identity is even higher as defined above.

In silico screening may comprise the steps of

i. performing a sequence similarity search of at least 10 nucleotides of the coding sequence SEQ ID No 1 (nucleotides no 52051 to 52466),

ii. aligning "hits" to said coding sequence,

iii. determining the presence of an open reading frame in the "hits".

One suitable method for performing the sequence similarity search is a Blast search with default parameters.

As for the in vitro method, the the open reading frame preferably encodes a polypeptide having at least 60 % sequence identity with the amino acid sequence of SEQ ID No 3. More preferably the sequence identify is even higher.

After having identified putative 4-helical cytokines, the function of these polypeptides may be assessed by synthesising the polypeptide encoded by the open reading frame and determining the activity of said polypeptide in a cytokine activity assay, preferably an interleukin assay, more preferably an interleukin-4 assay. Having verified the function of the 4-helical cytokine, it is also within the scope of the present invention to further formulate the polypeptide with a pharmaceutically acceptable carrier or diluent and obtain a pharmaceutical composition.

Hybridisation

The entire nucleotide sequence of the coding sequence of SEQ ID No 1 or portions thereof can be used as a probe capable of specifically hybridising to corresponding sequences. To achieve specific hybridisation under a variety of conditions, such probes include sequences that are unique and are preferably at least about 10 nucleotides in length, and most preferably at least about 20 nucleotides in length. Such probes can be used to amplify corresponding sequences from a chosen organism or subject by the well-known process of polymerase chain reaction (PCR) or other amplification techniques. This technique can be used to isolate additional nucleotide sequences from a desired organism or as a diagnostic assay to determine the presence of the coding sequence in an organism or subject. Examples include hybridisation screening of plated DNA libraries (either plaques or

colonies; see e. g. Innis et al. (1990) PCR Protocols, A Guide to Methods and Applications, eds., Academic Press).

The terms "stringent conditions" or "stringent hybridisation conditions" include reference to conditions under which a probe will hybridise to its target sequence, to a detectably greater degree than other sequences (e. g., at least twofold over background). Stringent conditions are target sequence dependent and will differ depending on the structure of the polynucleotide. By controlling the stringency of the hybridisation and/or washing conditions, target sequences can be identified which are 100% complementary to a probe (homologous probing).

Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing).

Generally, probes for hybridisation of this type are in a range of about 1000 nucleotides in length to about 250 nucleotides in length.

An extensive guide to the hybridisation of nucleic acids is found in Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, Part I, Chapter 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York (1993); and Current Protocols in Molecular Biology, Chapter 2, Ausubel, et al., Eds., Greene Publishing and Wiley-Interscience, New York (1995). See also Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual (2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.).

Specificity is typically the function of post-hybridisation washes, the critical factors being the ionic strength and temperature of the final wash solution.

Generally, stringent wash temperature conditions are selected to be about 5°C to about 2°C lower than the melting point (Tm) for the specific sequence at a defined ionic strength and pH. The melting point, or denaturation, of DNA occurs over a narrow temperature range and represents the disruption of the double helix into its complementary single strands. The process is described by the temperature of the midpoint of transition, Tm, which is also called the melting temperature.

Formulas are available in the art for the determination of melting temperatures.

Preferred hybridisation conditions for the nucleotide sequence of the invention include hybridisation at 42°C in 50% (w/v) formamide, 6X SSC, 0.5% (w/v) SDS, 100 mg/ml salmon sperm DNA. Exemplary low stringency washing conditions include hybridization at 42°C in a solution of 2X SSC, 0.5% (w/v) SDS for 30 minutes and repeating. Exemplary moderate stringency conditions include a wash in 2X SSC, 0.5% (w/v) SDS at 50°C for 30 minutes and repeating.

Exemplary high stringency conditions include a wash in 2X SSC, 0.5% (w/v) SDS, at 65°C for 30 minutes and repeating. Sequences that correspond to the AMB-1 gene or fractions thereof according to the present invention may be obtained using all the above conditions. For purposes of defining the invention, the high stringency conditions are used.

Promoters, Enhancers, and Signal Sequence Elements

The promoters and enhancers that control the transcription of protein-encoding genes are composed of multiple genetic elements. The cellular machinery is able to gather and integrate the regulatory information conveyed by each element, allowing different genes to evolve distinct, often complex patterns of transcriptional regulation.

The term promoter will be used here to refer to a group of transcriptional control modules that are clustered around the initiation site for RNA polymerase II. Much of the thinking about how promoters are organized derives from analyses of several viral promoters, including those for the HSV thymidine kinase (tk) and SV40 early transcription units. These studies, augmented by more recent work, have shown that promoters are composed of discrete functional modules, each consisting of approximately 7-20 bp of DNA, and containing one or more recognition sites for transcriptional activator proteins. At least one module in each promoter functions to position the start site for RNA synthesis. The best known example of this is the TATA box, but in some promoters lacking a TATA box, such as the promoter for the mammalian terminal deoxynucleotidyl transferase gene and the promoter for the SV

40 late genes, a discrete element overlying the start site itself helps to fix the place of initiation.

Additional promoter elements regulate the frequency of transcriptional initiation. Typically, these are located in the region 30-110 bp upstream of the start site, although a number of promoters have recently been shown to contain functional elements downstream of the start site as well. The spacing between elements is flexible, so that promoter function is preserved when elements are inverted or moved relative to one another. In the tk promoter, the spacing between elements can be increased to 50 bp apart before activity begins to decline. Depending on the promoter, it appears that individual elements can function either cooperatively or independently to activate transcription.

Enhancers were originally detected as genetic elements that increased transcription from a promoter located at a distant position on the same molecule of DNA. This ability to act over a large distance had little precedent in classic studies of prokaryotic transcriptional regulation.

Subsequent work showed that regions of DNA with enhancer activity are organized much like promoters. That is, they are composed of many individual elements, each of which binds to one or more transcriptional proteins.

The basic distinction between enhancers and promoters is operational. An enhancer region as a whole must be able to stimulate transcription at a distance; this need not be true of a promoter region or its component elements. On the other hand, a promoter must have one or more elements that direct initiation of RNA synthesis at a particular site and in a particular orientation, whereas enhancers lack these specificities. Aside from this operational distinction, enhancers and promoters are very similar entities. They have the same general function of activating transcription in the cell. They are often overlapping and contiguous, often seeming to have a very similar modular organisation. Taken together, these considerations suggest that enhancers and promoters are homologous entities and that the transcriptional activator proteins bound to these sequences may interact with the cellular transcriptional machinery in fundamentally the same way.

Polynucleotide variants

The following terms are used to describe the sequence relationships between two or more polynucleotides: "predetermined sequence", "comparison window", "sequence identity", "percentage of sequence identity", and "substantial identity".

A "predetermined sequence" is a defined sequence used as a basis for a sequence comparision; a predetermined sequence may be a subset of a larger sequence, for example, as a segment of a full-length DNA or gene sequence given in a sequence listing, such as a polynucleotide sequence of SEQ ID No. 1 or 2, or may comprise a complete DNA or gene sequence. Generally, a predetermined sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length.

Since two polynucleotides may each (1) comprise a sequence (i.e., a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window", as used herein, refers to a conceptual segment of at least 20 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a predetermined sequence of at least 20 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the predetermined sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences.

Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2: 482, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48: 443, by the search for similarity method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. (U.S.A.) 85: 2444, by computerised implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575

Science Dr., Madison, Wis.), or by inspection, and the best alignment (i.e., resulting in the highest percentage of homology over the comparison window) generated by the various methods is selected.

The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparision (i.e., the window size). The terms "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 85 percent sequence identity, preferably at least 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a predetermined sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 25-50 nucleotides, wherein the percentage of sequence identity is calculated by comparing the predetermined sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of the predetermined sequence over the window of comparison. The predetermined sequence may be a subset of a larger sequence, for example, as a segment of the full-length SEQ ID No. 1 polynucleotide sequence illustrated herein.

Site-Specific Mutagenesis

Site-specific mutagenesis is a technique useful in the preparation of individual peptides, or biologically ***functional equivalent*** proteins or peptides, through specific mutagenesis of the underlying DNA. The technique, well-known to those of skill in the art, further provides a ready ability to prepare and test sequence variants, for example, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the DNA. Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient

size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Typically, a primer of about 14 to about 25 nucleotides in length is preferred, with about to about 10 residues on both sides of the junction of the sequence being altered.

In general, the technique of site-specific mutagenesis is well known in the art, as exemplified by various publications. As will be appreciated, the technique typically employs a phage vector which exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage are readily commercially-available and their use is generally well-known to those skilled in the art. Double-stranded plasmids are also routinely employed in site directed mutagenesis which eliminates the step of transferring the gene of interest from a plasmid to a phage.

In general, site-directed mutagenesis in accordance herewith is performed by first obtaining a single-stranded vector or melting apart of two strands of a double-stranded vector which includes within its sequence a DNA sequence which encodes the desired peptide. An oligonucleotide primer bearing the desired mutated sequence is prepared, generally synthetically. This primer is then annealed with the single-stranded vector, and subjected to DNA polymerizing enzymes such as E. coli polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as E. coli cells, and clones are selected which include recombinant vectors bearing the mutated sequence arrangement.

Mutagenesis of a preferred predetermined fragment of 4-helical cytokine can be conducted by making amino acid insertions, usually on the order of about from 1 to 10 amino acid residues, preferably from about 1 to 5 amino acid residues, or deletions of from about from 1 to 10 residues, such as from about 2 to 5 residues.

The preparation of sequence variants of the selected peptide-encoding DNA segments using site-directed mutagenesis is provided as a means of producing potentially useful species and is not meant to be limiting as there are other ways in

which sequence variants of peptides and the DNA sequences encoding them may be obtained. For example, recombinant vectors encoding the desired peptide sequence may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants. Specific details regarding these methods and protocols are found in the teachings of Maloy et al. (1994); Segal (1976); Prokop and Bajpai (1991); and Maniatis et al.(1982), each incorporated herein by reference, for that purpose.

The PCR-based strand overlap extension (SOE) for site-directed mutagenesis is particularly preferred for site-directed mutagenesis of the nucleic acid compositions of the present invention. The techniques of PCR are well-known to those of skill in the art, as described hereinabove. The SOE procedure involves a two-step PCR protocol, in which a complementary pair of internal primers (B and C) are used to introduce the appropriate nucleotide changes into the wild-type sequence. In two separate reactions, flanking PCR primer A (restriction site incorporated into the oligo) and primer D (restriction site incorporated into the oligo) are used in conjunction with primers B and C, respectively to generate PCR products AB and CD. The PCR products are purified by agarose gel electrophoresis and the two overlapping PCR fragments AB and CD are combined with flanking primers A and D and used in a second PCR reaction. The amplified PCR product is agarose gel purified, digested with the appropriate enzymes, ligated into an expression vector, and transformed into E. coli JM101, XL1-Blue® (Stratagene, La Jolla, Calif.), JM105, TG1 (Carter et al., 1985), or other such suitable cells as deemed appropriate depending upon the particular application of the invention. Clones are isolated and the mutations are confirmed by sequencing of the isolated plasmids. Beginning with the native gene sequences, for example, the nucleic acid sequences encoding eukaryotic disulfide-bond-containing polypeptides such as PTI or tPA and the like, suitable clones and subclones may be made in the appropriate vectors from which site-specific mutagenesis may be performed.

4-helical cytokine receptors and binding modulators

Receptors for the 4-helical cytokines may be identified by contacting the isolated polypeptide or an expression vector encoding said isolated polypeptide with at least one cell line being dependent on a specific cytokine and observing at least one

parameter selected from the group consisting of: proliferation, apoptosis, necrosis, cell cycle changes or other physiological responses, inhibition of /activation of enzymes or caspases, upregulation of/ degradation of mRNA or proteins involved in proliferation, apoptosis, necrosis or cell cycle changes. By comparing the response with the response of the cell line to known cytokines and in particular to known interleukins, the receptor for the novel 4-helical cytokines can be identified.

An alternative method comprises the steps of contacting the isolated polypeptide with a plurality of putative polypeptides and selecting polypeptides that bind to the isolated polypeptide as receptors. This can conveniently be done by binding the 4-helical cytokines to a solid surface, or by binding the plurality of polypeptides are to a solid surface.

The K_D between the receptor and the isolated polypeptide is preferably less than 500 µM, more preferably less than 250 µM, more preferably less than 100 µM, more preferably less than 10 µM, more preferably less than 1 µM, more preferably less than 100 nM, more preferably less than 10 nM, such as less than 1 nM, for example less than 100 pM, such as less than 10 pM, for example less than 1 pM.

In order to identify a specific receptor for the novel class of cytokines or for a single member of this class the method further comprises selecting those receptors that bind the isolated polypeptide with higher affinity than they bind IL4, IL13, IL3, GM-CSF.

Having identified novel 4-helical cytokines and their receptors it is also within the scope of the present invention to identify a modulator of the binding between an isolated polypeptide according to the invention and a receptor identified according to the invention. The method comprises providing a complex between said polypeptide and said receptor, said complex having a predetermined K_D, and providing a plurality of putative modulators, contacting said complex with said plurality of putative modulators, and selecting those modulators that cause an increase in the K_D of at least 10%, more preferably more than 20 %, more preferably more than 50 %, more preferably more than 100 %, more preferably more than 200 %, more preferably more than 5 times, more preferably more than 10 times, such as more

than 100 times, for example more than 1000 times, such as more than 10,000 times, for example more than 100,000 times, such as more than 1,000,000 times.

Drug screening

The invention provides different methods for identifying compounds capable of treating B-CLL. In a first embodiment the method comprises administering a test-compound to a host cell comprising a recombinant expression construct, said expression construct comprising the promoter sequence of bases no. 40001 to 51417 or 40001 to 49100 of SEQ ID No 1 or a fragment thereof operably linked to a reporter gene, and determining the presence and/or amount of the reporter gene product. This method specifically addresses compounds capable of regulating transcription from the AMB-1 gene.

Suitable reporter genes are selected from the group consisting of encoding a coloured product, such as green fluorescent protein, GUS, luciferase, an apoptotic product, lux gene, CAT (chloramphenicol acetyl transferase).

Another method for screening for a compound capable of treating B-CLL, comprises administering a test-compound to a host cell comprising a recombinant expression construct, said expression construct comprising a constitutive promoter directing the expression of a polypeptide according to the invention and on said cell measuring a parameter selected from the group consisting of: proliferation, apoptosis, necrosis, cell cycle changes or other physiological responses, inhibition of /activation of enzymes or caspases, upregulation of/ degradation of mRNA or proteins involved in proliferation, apoptosis, necrosis or cell cycle changes.

Preferably the host is a non-human mammal, such as a rodent such as mouse or rat. Whole animals may be used for the biological assays, in particular rodents.

A further method for screening for a compound capable of treating B-CLL, comprising administering a test-compound to a cell line established from a subject diagnosed according to the invention, said method comprising measuring in said cell line: proliferation, apoptosis, necrosis, cell cycle changes or other physiological responses, inhibition of /activation of enzymes or caspases, upregulation of/

degradation of mRNA or proteins involved in proliferation, apoptosis, necrosis or cell cycle changes.

Mutations

Finally, the invention provides a method for determining an increased or decreased predisposition for B-CLL comprising determining in a biological sample from a subject a germline alteration in a target nucleic acid sequence comprising 150,000 nucleotides, said target nucleic acid sequence comprising at least 10 nucleotides of SEQ ID No 1. This aspect is based on the finding of the importance of the expression product of SEQ ID No 1, and the complete absence of any detectable expression product of SEQ ID No 1 in healthy tissue and in patients with good prognosis B-CLL. It is highly likely that the difference is caused by a germline alteration. A germline alteration can be targeted by gene therapy methods and by the methods provided in the present invention.

Preferably, said predisposition is a predisposition for poor prognosis of B-CLL.

Examples

Example 1: Bioinformatic analysis of AMB-1.

The two AMB1 cDNAs, AMB1-short and AMB1-long, comprises 3893 and 6209 nucleotides, respectively. The largest coding sequence is from pos. 3001 to 3363 (stop codon 3364-3366) in AMB1-long and 685 to 1047 (stop codon 1048-1050) in the AMB1-short. The open reading frame encodes a peptide of 121 amino acids. Comparison with the genomic sequence on chromosome 12 has revealed that the cDNA is derived from two exons, exon 1 of 4254 (AMB-long) or 1938 (AMB1-short) nucleotides and exon 2 of 1955 nucleotides (both long and short form), separated by an intron of 3099 nucleotides.

The DNA and protein sequence data bases (GenBank and EBI) have been searched for sequences with similarity to AMB1. The only significant match to the complete mRNA sequence and the DNA sequence of the putative coding region were BAC clones derived from the region on human chromosome 12 where the gene is located. Searches with the peptide sequence in the sptmr data base of

peptide sequences (includes Sprot and nrtrembl) showed a low similarity to putative intron maturases from cloroplasts and to bovine IL4 (Fig. 4). The percentage similarity to both maturases and bovine IL4 was low (25.6% and 30.3%, respectively) and the similarity to maturases only included a match to 75 amino acids of the much larger maturases. In contrast, the match to bovine IL4 extended over the full peptide sequence. IL4, and other 4-helical cytokines, include a leader peptide sequence (signal peptide) allowing the proteins to be secreted. The AMB1 peptide sequence includes a N-terminal peptide sequence with similarity to signal peptide sequences, however, it is not a typical sequence.

A 3D search has been performed, where a peptide sequence is searched for similarity to known protein or peptide 3D-structures. The two best matches were the thioredoxin fold and the human 4-helical cytokine IL4 (Fig. 5). The two matches had almost similar probability scores (2.88 and 3.05, respectively). Searches with 4-helical cytokine peptide sequences (IL4, IL3, IL13 and GM-CSF) revealed that all could be folded into both a 4-helical cytokine structure and the thioredoxin fold. Thus, the AMB1 peptide sequence share this property with 4-helical cytokines. The structural similarity is not perfect (Fig. 6) and there are no obvious glycosylation sites in the AMB1 sequence, however, the similarity is significant. Alignment of the AMB1 peptide sequence with the sequences of IL4, IL3, IL13 and GM-CSF, based on their structures, showed very litle sequence conservation but a high degree of structural conservation (Fig. 7). Based on this alignment, AMB1 has similarities to all the 4-helical cytokines, and the length of AMB1 and the position of gaps in the alignment could suggest a higher similarity to eg. IL13, but searches at 3D-PSSM only identified a significant similarity to the structure of IL4, not IL13, IL3 or GM-CSF. However, the search algorithms are not perfect and may therefore not detect a possible low structural similarity.

Example 2: Differential expression of AMB-1

Patient material

Blood samples were collected from newly diagnosed untreated patients with B-CLL. Mononuclear cells were isolated by Lymphoprep separation (Nycomed Pharma, Oslo, Norway), and the percentage of CD5+CD20+ B-CLL cells in the mononuclear fraction was >90% in all samples as determined by flow cytometric analysis.

Isolation of RNA and conversion to cDNA.

Material for RNA production was isolated mononuclear cells from B-CLL patients or mononuclear cells from lymphoprep separated buffy coats from normal donors. Total RNA was isolated from $5x10^7$ or more cells using the QIAamp RNA Blood Mini kit (Qiagen, Valencia, CA) with DNAse treatment. RNA (1ug) was converted to cDNA by incubation with a mixture of random-primers (1µg) and T24-primer (1µg) for 5 minutes at 70°C. After cooling on ice, the reaction mixture was added to a final volume of 25µl containing 30U of AMV Reverse Transcriptase HC (Promega, Madison, WI, USA), 1x First Strand Buffer (50mM Tris-HCl, pH 8.3, 50mM KCl, 10mM $MgCl_2$, 10mM DTT, 0.5mM spermidine), 2.5mM of each dNTP and 60U rRNasin ribonuclease inhibitor (Promega, Madison, WI, USA). The reaction was performed for 60 minutes at 37°C.

Determination of somatic hypermutation status

Two µl of cDNA was amplified using a GeneAmp PCR System 2700 (Applied Biosystems, Warrington, UK) with a 40 pmol specific upstream primer corresponding to 1 of the 6 human VH family leader sequences (VH1: 5'-CCATGGACTGGACCTGGAGG-3', VH2: 5'-ATGGACATACTTTGTTCCAGC-3', VH3: 5'-CCATGGAGTTTGGGCTGAGC-3', VH4: 5'-ATGAAACACCTGTGGTTCTT-3', VH5: 5'-ATGGGGTCAACCGCGATCCT-3', VH6: 5'-ATGTCTGTCTCCTTCCTCAT-3') and a 40 pmol downstream primer (Cµ:5'-GAGGCTCAGCGGGAAGACCTT-3' or Cγ:5'-GGGGAAGACCGATGGGCCCCT-3') corresponding to a consensus sequence of the constant region of IgM or IgG respectively. The Reverse Transcription (RT)-PCR reaction contained 1xPCR buffer (10mM Tris-HCl, pH 9.0, 50mM KCl, 0.1% Triton X-100), 2.5mM $MgCl_2$, 0.2mM of each dNTP and 1.5U Taq DNA Polymerase (Promega, Madison, WI, USA) in a final volume of 100µl. The RT-PCR was performed under the following conditions: 1 cycle of 94°C for 5 minutes, 30 cycles of denaturation at 94°C for 30 secs, annealing at 62°C for 30 sec. and extension at 72°C for 30 sec, and a final extension at 72°C for 7 minutes. The RT-PCR products were analysed on 2% agarose gels and sequenced in an HBI Prism 310 Genetic Analyzer (Perkin Elmer, Foster City ,CA, USA) using the BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Warrington, UK) following the manufacturer's instructions.

Sequences obtained from each sample were compared to germ line sequences in the V base sequence directory (I.M. Tomlinson, MRC Center for Protein Engineering, Cambridge, UK) using BLAST, and the closest germ line sequence was assigned. A gene sequence was considered to be mutated if it had equal or more than 2% sequence alterations when compared to the closest published germ line sequence.

mRNA isolation

The full length AMB1 mRNA was isolated from unmutated patients by the RACE-PCR (rapid amplification of cDNA ends-polymerase chain reaction) approach using the SMART RACE cDNA amplification kit (Clontech, Palo Alto, CA) according to the manufacturer's instructions. The antisense primer sequence was 5'-TACATTACCAACACACGCGCAACAG-3'.

RT-PCR

To evaluate the mRNA expression pattern of AMB1 in unmutated and mutated B-CLL patients RT-PCR was performed. Exon-overlapping oligonucleotide primers were: 5'-ATCCAGCCAGGATGAAATAGAA-3' and 5'-CACTTGTCACACACATAAAGG-3'. The RT-PCR was performed in a GeneAmp PCR System 2700 thermal cycler with an initial denaturation at 94°C for 2 minutes, 40 cycles of 96°C for 25 sec., 62°C for 25 sec. and 72°C for 90 secs, and a final extension at 72°C for 5 minutes. The reactions contained 2µl cDNA, 1x DDRT-PCR buffer (10mM Tris-HCl, pH 8.3, 50mM KCl, 1.8mM $MgCl_2$, 0.1% Triton X-100, 0.005% gelatine), 0.25mM of each dNTP, 30 pmol of each primer and 0.5U Taq DNA Polymerase (Promega, Madison, WI, USA) in a 30µl final volume. RT-PCR products were analyzed by gelelectrophoresis on 2% agarose gels and visualized with a Gene Genius Bio Imaging System (Syngene, Frederick, MD) after staining with ethidium bromide.

Statistical analysis

Statistical significance of the correlation between somatic hypermutation status and AMB1 expression was analyzed using Fisher's exact test.

Northern blotting. RNA from spleen, bone marrow and colon was purchased from Clontech. The AMB1 probe was an 896 base pair fragment (57661-56766) obtained

by RT-PCR as described above with the primers 5'-TCACCTGGGAGCTCAGAGGA-3' and 5'-GTGATCCTGGGAGAATCTCT-3'. For Northern blotting, 5 µg of RNA was run on a 1% agarose-gel with 6% formaldehyde dissolved in 1 x MOPS (20 mM 3-(N-morpholino)- propane-sulfonic acid, 5 mM sodium acetate, 1 mM EDTA, pH 7.0) for size separation. The presence of equal amounts of RNA in each lane was ensured by ethidium bromide staining. The RNA was transferred to a Hybond-N membrane (Amersham, Little Chalfont, UK) by capillary blotting and fixed by UV-irradiation. The filters were pre-hybridized for 1-2 hours at 42°C in 6 ml ULTRAhyb (Ambion, Austin, TX, USA) preheated to 68°C and hybridized overnight at 42°C after addition of further 4 ml containing the ^{32}P-labeled probe and sheared salmon sperm DNA (10 µg/ml). The membranes were washed for 2 x 15 min. at 42°C in 2 x SSC, 0.1% SDS followed by 1 x 15 min. in 0.2 x SSC, 0.1 % SDS and 2 x 15 min. in 0.1 x SSC, 0.1 % SDS at 42°C. The blot was developed and quantified by a phosphoimager. The sizes of the mRNAs were determined by reference to 18S and 28S ribosomal RNA, which were visualized by ethidium bromide staining. The AMB1 probe used for hybridization was radiolabeled with [α-^{32}P] dCTP using the Random Primers DNA Labeling System (Gibco BRL).

Dot blot of multiple tissue expression (MTE) array. An MTE array (Clontech, Palo Alto, CA, USA) was hybridised to AMB1 at 65°C in ExpressHyb (Clontech) supplemented with sheared salmon sperm DNA (7.5 µg/ml) and human C_ot-1 DNA (1.5 µg/ml) according to the manufacturers recommendations. The tissue types represented on the MTE array are shown in Figure 11. Following hybridisation the filter was washed 5 x 20 min. at 65°C in 2 x SSC (1 x SSC =150 mM NaCl, 15 mM sodium citrate, pH 7.0), 1% SDS and 2 x 20 min at 65°C in 0.1 x SSC, 0.5% SDS. The blot was developed and quantified by a phosphoimager (Fuji Imager Analyzer BAS-2500, Image Reader ver. 1.4E, Image Gauge ver. 3.01 software, Fuji, Stockholm, Sweden). The membranes were stripped by boiling in 0.5% SDS for 10 min. before rehybridization. The probe used for hybridization were radiolabeled with [α-^{32}P] dCTP using the Random Primers DNA Labeling System (Gibco BRL, Rockville, ML, USA).

Results

Blood samples were collected for the B-CLL patient database from newly diagnosed, untreated B-CLL patients. The degree of somatic hypermutation was determined by sequencing of the Ig VH region and alignments to BLAST or DNAPlot databases, with a cut-off level for Ig VH homology to the nearest germ line sequence of 98%. By DDRT-PCR a gene (hereafter referred to as AMB-1) was found that is expressed in unmutated patients with poor prognosis. This gene is not found in the mutated patients. When AMB-1 was sequenced and aligned to known sequences in GenBank, perfect homology was found to 225 base pairs (bp) of human genomic DNA from chromosome 12. Importantly, aberrations at chromosome 12 are among the most frequent cytogenetic abnormalities in B-CLL, and moreover, AMB-1 mapped to a region on chromosome 12 that is known to harbor molecular aberrations in B-CLL. The "AMB-1 gene" had not been annotated as a gene on the chromosome.

Since the 225 bp gene sequence found by DDRT-PCR aligned perfectly to genomic DNA sequence on chromosome 12, it has been possible to use PCR and RACE analysis to identify more of the upstream AMB-1 sequence. At present 6209 bp of a mRNA has been identified. This mRNA consists of two exons separated by a 3o99 bp intron. An open reading frame is present at pos. 3001 to 3363 encoding a protein of 121 amino acids. There is no significant DNA sequence similarity to any known gene. In particular, the coding region of the AMB1 mRNA is not present in any known EST. The protein with the highest similarity to the AMB1 protein sequence, was bovine IL4 (30%). Based on the known sequence of AMB-1 an RT-PCR with primers that extend across the intron was set up. As shown in Figure 2, the RT-PCR confirmed that AMB-1 is expressed in the unmutated patients (UPN1-8) while no expression of AMB-1 is seen in mutated patients (UPN9-16).

Northern blot analysis was performed to determine the size of AMB-1's mRNA transcript. As shown in Figure 3 the probe identifies predominantly a transcript of about 4000 bp, but also a smaller and a very large transcript from the three patients without somatic hypermutation (UPN1, UPN4 and UPN7). However, the probe does not recognise any transcripts from the patients with somatic hypermutation (UPN9, UPN10, UPN13, UPN21) or the various cell lines and tissue samples. Similar results were obtained when cell lines and tissue samples were investigated for the presence of AMB-1 by RT-PCR (results not shown). Dot blot analysis on a

purchased filter with 96 different RNA samples (Figure 11) only revealed specific binding to the total DNA control dot, but not to any specific tissue. A fragment of 225 C-terminal basepairs of the AMB-1 mRNA was screened against human cDNA libraries from normal tissues and cell lines including foetal tissue (see table below), but the AMB-1 mRNA was not present in any of these libraries. This strengthens that AMB-1 is only expressed in unmutated CLL. Thus AMB-1 is only expressed in B-CLL cells without hypermutation or AMB-1 is expressed at extremely low levels in other tissues.

1	Adult Brain (1-3 kb)	13	Adult Brain (>3 kb)	25	Placenta	37	Amygdala
2	Spleen (1-3 kb)	14	Spleen (>3 kb)	26	Stomach	38	Corpus Callo-sum
3	Liver (1-3 kb)	15	Liver (>3 kb)	27	Mammary	39	Adult brain-2
4	Heart (1-3 kb)	16	Heart (>3 kb)	28	Prostate	40	Foetal brain-2
5	Spinal Cord (1-3 kb)	17	Spinal Cord (>3 kb)	29	Pancreas		
6	Small Intestine (1-3 kb)	18	Small Intestine (>3 kb)	30	Sunstantia Nigra		PBL (separate screen)*
7	Colon (1-3 kb)	19	Colon (>3 kb)	31	Foetal Brain		
8	Skeletal Muscle (1-3 kb)	20	Skeletal Muscle (>3 kb)	32	Pituitary		
9	Bone Marrow (1-3 kb)	21	Bone Marrow (>3 kb)	33	Caudate Nucleus		
10	Kidney (1-3 kb)	22	Kidney (>3 kb)	34	Cerebellum		
11	Lung (1-3 kb)	23	Lung (>3 kb)	35	Thalamus		
12	Testis (1-3 kb)	24	Testis (>3 kb)	36	Hippocampus		

*Peripheral blood lymphocyte

We next tested the predictive value, in terms of Ig VH mutational status, of expression of AMB-1 in 29 consecutive newly diagnosed patients. At present 13 somatically unmutated and 16 somatically mutated patients have been included in our prospective patient database. The sensitivity and specificity for expression of AMB-1 in predicting mutational status is well above 90% ($p<0.0001$), which is at the level obtained by sequencing.

Example 3. Investigation of the prognostic significance of AMB-1 in terms of patient survival

Rationale: AMB-1 can be used to distinguish between the unmutated and mutated B-CLL patients. To get a better understanding of the prognostic value of AMB-1 the expression of AMB-1 is analysed by RT-PCR in patient samples from The Danish CLL-2 study. The Danish CLL-2 study (headed by Christian Geisler and Mogens Mørk Hansen) has accrued 549 consecutive and newly diagnosed B-CLL patients

from 1982 to 1984. The study comprises one of the largest prospective studies of B-CLL prognosis, including analysis of clinical stage, response to therapy, bone marrow infiltration pattern, immunophenotype and cytogenetics. The median follow-up time from this study is now above 25 years.

Methods: The sample material consists of procured smears and frozen samples from the CLL-2 study. RNA is extracted from the stored smears and cDNA is made. First, RT-PCR is performed on the samples using primers that extend across the intron to avoid inconsistencies from possible DNA contamination in the samples. The ability of AMB-1 expression to predict mutational status, chromosomal aberrations and overall survival will be tested in multivariate analysis.

Second, a Real Time PCR analysis is established, based on Taqman technology, in order to analyze the importance of quantitative expression of AMB-1.

Third, in-situ hybridization is used to determine if AMB-1 is globally expressed, or only expressed in a fraction of the malignant population of cells.

Example 4: Identification of possible cytogenetic aberrations near or within the region encoding AMB-1 on chromosome 12.

Rationale: The limited expression profile of AMB-1 suggests that it may be a result of a genetic aberration (e.g. deletion, translocation or alternative splicing) or that the promotor region controlling the expression of AMB-1 is uniquely activated in unmutated B-CLL. Another gene is situated about 200.000 bases upstream of the AMB-1 gene (SEQ ID No 1) on chromosome 12 and the inventors we have determined that this gene is expressed at equal levels in unmutated and mutated patients.

Methods: Using primers, initially spaced about 20.000 bp apart; this region on chromosome 12 is characterised in unmutated B-CLL patients. If genetic aberrations within the region are detected by PCR analysis of chromosomal DNA, detailed molecular genetic studies using FISH, microsatellite analysis and Southern blotting will be employed. The whole region from unmutated patients is sequenced.

Example 5: Assay for the biological activity of 4-helix cytokines.

The assay is based on the use of a cytokine dependent or stimulated cell line, for example an IL4 dependent cell line ("Optimisation of the CT.h4S bioassay for detection of human interleukin-4 secreted by mononuclear cells stimulated by phytohaemaglutinin or by human leukocyte antigen mismatched mixed lymphocyte culture", Petersen, S.L., Russell, C.A., Bendtzen, K. & Vindeløv, L.L., Immunology Letters 84 (2002) 29-39). Other examples of cytokine dependent cell lines include IL13 dependent cell lines. A list of commercially available cytokine dependent cell lines is disclosed in the general part of the description. These can all be used for assessing cytokine activity. The most preferred cell lines are those that are IL4 dependent.

The assay can be performed in two ways. The first assay comprises providing recombinantly produced AMB1 protein or a functional equivalent thereof and determine the proliferation rate of the cell line. The proliferation rate (either rate of proliferation or ± proliferation) can be compared to the proliferation rate of the cell line exposed to IL4 or another known 4-helical cytokine or interleukin.

If a positive result is obtained with a polypeptide an assay will be performed on the same cell line with the IL4 receptor blocked. This will check whether the stimulus goes through IL4R.

The second assay is based on transfection of a gene encoding a 4-helix cytokine according to the invention into cytokine dependent cells and observe proliferation or non-proliferation during transient expression.

Example 6: Cytokine receptor binding assays

The following is a description of the layout of a cytokine receptor binding assay used to determine the cytokine activity of the 4-helix cytokines according to the present invention.

The assays can be performed with any cytokine receptor. Preferred receptors include but is not limited to the receptors for IL4 IL13, IL3, and GM-CSF.

The ability of recombinant cytokine receptor to bind to 4-helical cytokine is assessed in a competitive binding ELISA assay as follows. Purified recombinant cytokine receptor (IL4, IL13, IL3 or GM-CSF receptors) (20 µg/ml in PBS) is bound to a Costar EIA/RIA 96 well microtiter dish (Costar Corp, Cambridge Mass., USA) in 50 µL overnight at room temperature. The wells are washed three times with 200 µL of PBS and the unbound sites blocked by the addition of 1% BSA in PBS (200 µl/well) for 1 hour at room temperature. The wells are washed as above. Biotinylated AMB-1 (1 µg/ml serially diluted in twofold steps to 15.6 ng/mL; 50 µL) is added to each well and incubated for 2.5 hours at room temperature. The wells are washed as above. The bound biotinylated AMB-1 is detected by the addition of 50 µl/well of a 1:2000 dilution of streptavidin-HRP (Pierce Chemical Co., Rockford, Ill.) for 30 minutes at room temperature. The wells are washed as above and 50 µL of ABTS (Zymed, Calif.) added and the developing blue color monitored at 405 nm after 30 min. The ability of unlabelled 4-helical cytokine to compete with biotinylated AMB-1, respectively, is assessed by mixing varying amounts of the competing protein with a quantity of biotinylated AMB-1 shown to be non-saturating (i.e., 70 ng/mL; 1.5 nM) and performing the binding assays as described above. A reduction in the signal (Abs 405 nm) expected for biotinylated 4-helical cytokine indicates a competition for binding to immobilised cytokine receptor.

The above identified assays can be used to identify 4-helical cytokines with similar binding affinities as AMB-1 (SEQ ID No. 3). In the competitive binding assays biotinylated IL4, IL13, IL3, or GM-CSF can be used to identify 4-helical cytokines which can compete with these cytokines.

Claims

1. A method for diagnosing a subtype of B-cell chronic lymphocytic leukaemia (B-CLL), said method comprising the steps of determining the presence or absence of a transcriptional or translational product of SEQ ID No 1 in a biological sample isolated from a subject.

2. The method of claim 1, wherein the B-CLL prognosis is poor.

3. The method of claim 1, wherein the subtype of B-CLL is characterised solely by the presence of a transcriptional or translational product of SEQ ID No 1.

4. The method of claim 1, wherein the subject is a mammal, preferably a human being.

5. The method of claim 4, wherein the mammal is selected from the group: domestic animals such as cow, horse, sheep, pig; and pets such as cat or dog.

6. The method of any of the preceding claims, wherein the transcriptional product is a mRNA sequence corresponding to SEQ ID No 2, SEQ ID No 4, or a fragment thereof.

7. The method of claim 6, wherein the presence or absence of the transcriptional product is determined by hybridisation techniques.

8. The method of claim 7, wherein the hybridisation is performed on a DNA array comprising an oligomer of at least 20 consecutive bases from the sequence 49101 – 53354 or 56454 – 58408 of SEQ ID No 1.

9. The method of claim 6, wherein the presence or absence of the transcriptional product is determined by specifically amplifying a transcriptional product having a sequence corresponding to SEQ ID No 2 or 4 or a fragment thereof.

10. The method of any of the preceding claims, wherein the translational product is a protein encoded by SEQ IN No 1 and/or 2 and/or 4.

11. The method of claim 10, wherein the detection is performed with an antibody directed against said protein, such as Western blotting, more preferably by using a fluorescently labelled antibody, preferably wherein the method comprises the use of FACS.

12. The method of claim 10, wherein detection of the protein comprises gel electrophoresis, gel filtration, ion exchange chromatography, FPLC.

13. The method of claim 10, wherein said protein is selected from the group comprising SEQ ID No 3 (protein), or a protein sharing at least 60 % sequence identity with SEQ ID No 3.

14. A method for determining the stage/progress of B-CLL comprising determining the amount of a transcriptional or translational product of SEQ ID No 1 in a biological sample isolated from a subject.

15. The method of claim 14, wherein the determination is performed during treatment to estimate the efficiency of such treatment.

16. The method according to any of the preceding claims, wherein the biological sample is selected from the group comprising: a blood sample, lymph node tissue, bone marrow, spinal liquid.

17. A method of treating B-CLL comprising administering to a subject being diagnosed according to any of the claims 1 to 13, a therapeutically effective amount of a compound capable of selectively killing and/or inhibiting division of and/or inducing apoptosis in B-CLL cells.

18. The method according to claim 17, wherein the compound is selected from the group chemotherapeutic agents, anti-CD20, or anti-CD52 or other antibodies, using non-myeloablative bone marrow transplantation.

19. A method of treating B-CLL comprising administering to a subject with a B-CLL diagnosis a compound capable of decreasing or inhibiting the formation of a transcriptional and/or translational product from SEQ ID No 1.

20. The method according to claim 19, wherein the compound is a therapeutic antibody directed against a polypeptide having the amino acid sequence of SEQ ID No 3, preferably wherein said antibody is a human or humanised antibody.

21. The method of claim 19, wherein the compound is an oligonucleotide capable of inhibiting transcription from SEQ ID No 1,

22. The method of claim 21, wherein said oligonucleotide comprises at least 8-10 consecutive nucleotides from the sequence 40001 to 51417 or the sequence 40001 to 49100 of SEQ ID No 1.

23. The method of claim 22, wherein said oligonucleotide comprises nucleotide monomers selected from the group: DNA, RNA, LNA, PNA, methylated DNA, methylated RNA, more preferably PNA or LNA.

24. The method of claim 21, wherein the compound is an oligonucleotide capable of binding to a transcriptional product and preventing translation, such as wherein the compound is an antisense construct or comprises a RNAi oligonucleotide.

25. The method according to claim 24, wherein the RNAi oligonucleotide comprises 8-22 consecutive nucleotides of the complementary sequence of SEQ ID No 2 and/or SEQ ID No 4, more preferably of SEQ ID No 2.

26. The method according to claim 25, wherein RNAi oligonucleotides are administered to the cell, or wherein a vector is transfected into the cells, said vector comprising a promoter region capable of directing the expression of at least one RNAi oligonucleotide, preferably wherein the cells comprise blood cells.

27. The method of claim 26, wherein said vector is coupled to a heparin receptor for targeting to blood cells.

28. The method according to claim 24, wherein the antisense construct comprises a promoter sequence capable of directing the transcription of at least part of the antisense equivalent of SEQ ID No 2 or 4.

29. The method of claim 26 or 28, wherein the antisense construct is targeted to B-CLL cells using the CD19 or CD20 receptor.

30. The method of claim 19, wherein the compound is a gene therapy vector comprising a promoter sequence operably linked to a sequence coding for a protein capable of inhibiting cell division in the cell and/or capable of killing the cell, said promoter sequence being a tissue specific promoter capable of directing expression only in B cells, more preferably only in B-CLL cells.

31. The method of claim 30, wherein said promoter sequence comprises bases No 40001 to 51417 of SEQ ID No 1 or a fragment thereof, such as the fragment from 40001 to 49100 or a fragment of this fragment.

32. The method of claim 31, wherein the promoter comprises at least 100 nucleotides 5' to base no. 51471 or 49100 of SEQ ID No 1, such as at least 200 nucleotides, for example at least 300 nucleotides, such as at least 400 nucleotides, for example at least 500 nucleotides, such as at least 600 nucleotides, for example at least 700 nucleotides, such as at least 800 nucleotides, for example at least 900 nucleotides, such as at least 1000 nucleotides, for example at least 1100 nucleotides, such as at least 1200 nucleotides, for example at least 1300 nucleotides, such as at least 1400 nucleotides, for example at least 1500 nucleotides, such as at least 1600 nucleotides, for example at least 1700 nucleotides, such as at least 1800 nucleotides, for example at least 1900 nucleotides, such as at least 2000 nucleotides, for example at least 2500 nucleotides, such as at least 3000 nucleotides, for example at least 3500 nucleotides, such as at least 5000 nucleotides, for example at least 10,000 nucleotides.

33. The method of claim 30, wherein said protein is selected from the group comprising: HSV-1 thymidine kinase, E. coli cytosine deaminase, the varicella-

zoster, virus thymidine kinase gene, the nitroreductase gene, the E. coli gpt gene, and the E. coli Deo gene.

34. A gene therapy vector capable of inhibiting or decreasing the formation of a transcriptional or translational product of SEQ ID No. 1.

35. The gene therapy vector of claim 34, comprising an oligonucleotide capable of inhibiting transcription from SEQ ID No 1.

36. The gene therapy vector of claim 35, wherein said oligonucleotide comprises at least 8-10 consecutive nucleotides from the sequence 40001 to 51417 or the sequence 40001 to 49100 of SEQ ID No 1.

37. The gene therapy vector of claim 36, wherein said oligonucleotide comprises nucleotide monomers selected from the group: DNA, RNA, LNA, PNA, methylated DNA, methylated RNA, more preferably PNA or LNA.

38. The gene therapy vector of claim 34, comprising an oligonucleotide capable of binding to a transcriptional product and preventing translation, such as wherein the compound is an antisense construct or comprises a RNAi oligonucleotide.

39. The gene therapy vector of claim 38, wherein the RNAi oligonucleotide comprises 8-24 consecutive nucleotides of the complementary sequence of SEQ ID No 2 and/or SEQ ID No 4, more preferably of SEQ ID No 2.

40. The gene therapy vector of claim 34, comprising a promoter capable of directing the transcription of a RNAi oligonucleotide comprises 8-24 consecutive nucleotides of SEQ ID No 2 and/or SEQ ID No 4, more preferably of SEQ ID No 2.

41. The gene therapy vector of claim 38, wherein the antisense construct comprises a promoter sequence capable of directing the transcription of at least part of the antisense equivalent of SEQ ID No 2 or 4.

42. The gene therapy vector of claim 40 or 41, wherein the promoter is a B-CLL specific promoter.

43. The gene therapy vector of claim 40 or 41, wherein the vector is targeted to B-CLL cells using a receptor selected from the group consisting of heparin, CD19, CD20.

44. The gene therapy vector of claim 34, comprising a promoter sequence operably linked to a sequence coding for a protein capable of inhibiting cell division in the cell and/or capable of killing the cell, said promoter sequence being a tissue specific promoter capable of directing expression only in B cells.

45. The gene therapy vector of claim 44, wherein said promoter sequence comprises bases No 40001 to 51417 of SEQ ID No 1 or a fragment thereof, such as the fragment no 40001 to 49100.

46. The gene therapy vector of claim 45, wherein the promoter comprises at least 100 nucleotides 5' to nucleotide no. 51418 or 49101 of SEQ ID No 1, such as at least 200 nucleotides, for example at least 300 nucleotides, such as at least 400 nucleotides, for example at least 500 nucleotides, such as at least 600 nucleotides, for example at least 700 nucleotides, such as at least 800 nucleotides, for example at least 900 nucleotides, such as at least 1000 nucleotides, for example at least 1100 nucleotides, such as at least 1200 nucleotides, for example at least 1300 nucleotides, such as at least 1400 nucleotides, for example at least 1500 nucleotides, such as at least 1600 nucleotides, for example at least 1700 nucleotides, such as at least 1800 nucleotides, for example at least 1900 nucleotides, such as at least 2000 nucleotides, for example at least 2500 nucleotides, such as at least 3000 nucleotides, for example at least 3500 nucleotides, such as at least 5000 nucleotides, for example at least 10,000 nucleotides.

47. The gene therapy vector of claim 44, wherein said protein is selected from the group comprising: HSV-1 thymidine kinase, E. coli cytosine deaminase, the varicella-zoster, virus thymidine kinase gene, the nitroreductase gene, the E. coli gpt gene, and the E. coli Deo gene.

48. An isolated polypeptide comprising or essentially consisting of the amino acid sequence of SEQ ID No. 3, or a fragment thereof, or a polypeptide functionally equivalent to SEQ ID No. 3, or a fragment thereof, wherein said fragment or functionally equivalent polypeptide has at least 60 % sequence identity with SEQ ID No 3 and

a) has interleukin or cytokine activity; and/or

b) is recognised by an antibody, or a binding fragment thereof, which is capable of recognising an epitope, wherein said epitope is comprised within a polypeptide having the amino acid sequence of SEQ ID No 3; and/or

c) is competing with a polypeptide having the amino acid sequence as shown in SEQ ID No 3 for binding to at least one predetermined binding partner.

49. The isolated polypeptide of claim 48, comprising or essentially consisting of the amino acid sequence of SEQ ID No. 3 or a fragment thereof.

50. The isolated polypeptide of claim 48, wherein the functionally equivalent polypeptide shares at least 60% sequence identity with SEQ ID No 3, more preferably at least 70% sequence identity, more preferably at least 80 % sequence identity, such as at least 90 % sequence identity, for example at least 95 % sequence identity, such as at least 97 % sequence identity, for example at least 98 % sequence identity.

51. The isolated polypeptide of claim 48, wherein the binding partner of item c) is selected from the group: an antibody directed against SEQ ID No 3, the receptor for IL4, IL3, IL13, GM-CSF, TGF-β, or IGF.

52. The isolated polypeptide of claim 48, which folds as a 4-helical cytokine.

53. The isolated polypeptide of claim 48, having interleukin activity, such as having IL3, IL13, GM-CSF, TGF-β, IGF activity, more preferably having IL4 activity.

54. A homo- or hetero-oligomer comprising at least one isolated polypeptides as defined in any of the claims 48 to 53, such as a dimer, a trimer, a quatramer, a quintamer, a hexamer, an octamer, a decamer, a dodecamer.

55. A pharmaceutical composition comprising an isolated polypeptide as defined in any of the claims 48 to 53 and a pharmaceutically acceptable carrier.

56. Use of an isolated polypeptide as defined in any of the claims 48 to 53 for the preparation of a medicament for the treatment of bone disorders, inflammation, for lowering blood serum cholesterol, allergy, infection, viral infections, hematopoietic disorders, preneoplastic lesions, immune related diseases, autoimmune related diseases, infectious diseases, tuberculosis, cancer, viral diseases, septic shock, reconstitution of the haematopoietic system, induction of the granulocyte system, pain, cardial dysfunction, CNS disorders, depression, artheritis, psoriasis, dermatitis, collitis, Chron's disease, diabetes, in a subject in need thereof.

57. Use of an isolated polypeptide as defined in any of the claims 48 to 53 as a growth factor.

58. Use of an isolated polypeptide as defined in any of the claims 48 to 53 as an adjuvant or as an immune anhancer.

59. Use of an isolated polypeptide as defined in any of the claims 48 to 53 for regulating TH2 immune responses.

60. Use of an isolated polypeptide as defined in any of the claims 48 to 53 for suppressing Th1 immune responses.

61. A method of vaccination against B-CLL said method comprising immunising a subject against a translational product of SEQ ID No 1.

62. The method of claim 61, comprising immunising said subject with at least one isolated polypeptide as defined in any of the claims 48 to 53 and optionally adjuvants and carriers.

63. The method of claim 61, comprising peptide loading of dendritic cells, or ex vivo expansion and activation of T-cells, or inducing a CTL response that targets cells expressing the polypeptide encoded by SEQ ID No 1.

64. A method for producing an antibody with specificity against an isolated polypeptide as defined in any of the claims 48 to 53, said method comprising the steps of

i) providing a host organism,

ii) immunising said host organism with an isolated polypeptide as defined in any of the claims 48 to 53, or transfecting said host organism with an expression vector capable of directing the expression of an isolated polypeptide as defined in any of the claims 48 to 53,

iii) obtaining said antibody.

65. The method of claim 64, wherein the host organism is a non-human mammal such as insect, preferably wherein the antibody is subsequently humanised.

66. The method of claim 64, further comprising formulating said antibody into a single-chain antibody.

67. The method of claim 64, wherein the host organism is a human being and the antibody is subsequently produced recombinantly in a non-human mammal, such as a mouse.

68. An antibody obtainable by the method of claim 64.

69. A pharmaceutical composition comprising an antibody according to claim 68.

70. The pharmaceutical composition according to claim 69 for treating cancer, preferably for treating leukaemia, more preferably for treating B-CLL leukaemia, more preferably for treating poor prognosis B-CLL leukaemia.

71. An expression vector encoding the antibody of claim 68.

72. An isolated polynucleotide selected from the group consisting of:

i) a polynucleotide comprising nucleotides 40001 to 60000 of SEQ ID No 1,

ii) a polynucleotide encoding a polypeptide having the amino acid sequence of SEQ ID No 3,

iii) a polynucleotide, the complementary strand of which hybridises, under stringent conditions, with a polynucleotide as defined in any of i) and ii), and encodes a polypeptide, which

a) has at least 60 % sequence identity with the amino acid sequence of SEQ ID No 3 and has interleukin or cytokine activity,

b) is recognised by an antibody, or a binding fragment thereof, which is capable of recognising an epitope, wherein said epitope is comprised within a polypeptide having the amino acid sequence of SEQ ID No 3; and/or

c) is competing with a polypeptide having the amino acid sequence as shown in SEQ ID No 3 for binding to at least one predetermined binding partner such as a cytokine receptor,

iv) a polynucleotide which is degenerate to the polynucleotide of iii), and

v) the complementary strand of any such polynucleotide.

73. The isolated polynucleotide according to claim 72, comprising the nucleotide sequence of SEQ ID No 2.

74. The isolated polynucleotide according to claim 72, comprising the nucleotide sequence of SEQ ID No 4.

75. A method for identifying a nucleotide sequence encoding a 4-helical cytokine, said method comprising the steps of:

i) isolating mRNA from a biological sample,

ii) hybridising the mRNA to a probe comprising at least 10 nucleotides of the coding sequence of SEQ ID No 1 (nucleotides no 52051 to 52466) under stringent conditions,

iii) determining the nucleotide sequence of a sequence capable of hybridising under step ii), and

iv) determining the presence of an open reading frame in the nucleotide sequence determined under step iii).

76. The method of claim 75, wherein the open reading frame encodes a polypeptide having at least 60 % sequence identity with the amino acid sequence of SEQ ID No 3.

77. A computer assisted method for identifying a nucleotide sequence encoding a 4-helical cytokine, said method comprising the steps of

i) performing a sequence similarity search of at least 10 nucleotides of the coding sequence SEQ ID No 1 (nucleotides no 52051 to 52466),

ii) aligning "hits" to said coding sequence,

iii) determining the presence of an open reading frame in the "hits".

78. The method of claim 76, wherein the sequence similarity search is a Blast search with default parameters.

79. The method of claim 76, wherein the open reading frame encodes a polypeptide having at least 60 % sequence identity with the amino acid sequence of SEQ ID No 3.

80. A method of preparing a 4-helical cytokine, said method comprising the steps of any of the claims 75 to 79, and further comprising synthesising the polypeptide encoded by the open reading frame and determining the activity of said polypeptide in a cytokine activity assay, preferably an interleukin assay, more preferably an interleukin-4 assay.

81. A method for preparing a pharmaceutical composition comprising the steps of claims 80 and further the step of formulating the polypeptide with a pharmaceutically acceptable carrier or diluent.

82. A method of identifying a receptor for an isolated polypeptide as defined in any of the claims 48 to 53, said method comprising the steps of contacting the isolated polypeptide or an expression vector encoding said isolated polypeptide with at least one cell line being dependent on a specific cytokine and observing at least one parameter selected from the group consisting of: proliferation, physiological response, cell cycle changes, apoptosis.

83. A method of identifying a receptor for an isolated polypeptide as defined in any of the claims 48 to 53, said method comprising the steps of contacting the isolated polypeptide with a plurality of polypeptides and selecting polypeptides that bind to the isolated polypeptide as receptors.

84. The method of claim 82, wherein the isolated polypeptide is immobilised by binding it to a solid surface, or wherein the plurality of polypeptides are immobilised by binding them to a solid surface.

85. The method of claim 82, wherein the K_D between the receptor and the isolated polypeptide is less than 500 µM, more preferably less than 250 µM, more preferably less than 100 µM, more preferably less than 10 µM, more preferably less than 1 µM, more preferably less than 100 nM, more preferably less than 10 nM, such as less than 1 nM, for example less than 100 pM, such as less than 10 pM, for example less than 1 pM.

86. The method of any of the claims 82 to 85, further comprising selecting those receptors that bind the isolated polypeptide with higher affinity than they bind IL4, IL13, IL3, GM-CSF.

87. A method for identifying a modulator of the binding between an isolated polypeptide according to any of the claims 48 to 53 and a receptor identified according to any of the claims 82 to 86, said method comprising providing a complex between said polypeptide and said receptor, said complex having a predetermined K_D, and providing a plurality of putative modulators, contacting said complex with said plurality of putative modulators, and selecting those modulators that cause an increase in the K_D of at least 10%, more preferably more than 20 %, more preferably more than 50 %, more preferably more than 100 %, more preferably more than 200 %, more preferably more than 5 times, more preferably more than 10 times, such as more than 100 times, for example more than 1000 times, such as more than 10,000 times, for example more than 100,000 times, such as more than 1,000,000 times.

88. A method for screening for a compound capable of treating B-CLL, comprising administering a test-compound to a host cell comprising a recombinant

expression construct, said expression construct comprising the promoter sequence of bases no. 40001 to 51417 or 40001 to 49100 of SEQ ID No 1 or a fragment thereof operably linked to a reporter gene, and determining the presence and/or amount of the reporter gene product.

89. The method of claim 88, wherein said host is a non-human mammal, such as a rodent such as mouse or rat.

90. The method of claim 88, wherein said reporter gene is selected from the group consisting of encoding a coloured product, such as green fluorescent protein, GUS, luciferase, an apoptotic product, lux gene, CAT (chloramphenicol acetyl transferase).

91. The method of claim 88, wherein the promoter comprises at least 100 nucleotides 5' to the transcription initiation site of SEQ ID No 1, such as at least 200 nucleotides, for example at least 300 nucleotides, such as at least 400 nucleotides, for example at least 500 nucleotides, such as at least 600 nucleotides, for example at least 700 nucleotides, such as at least 800 nucleotides, for example at least 900 nucleotides, such as at least 1000 nucleotides, for example at least 1100 nucleotides, such as at least 1200 nucleotides, for example at least 1300 nucleotides, such as at least 1400 nucleotides, for example at least 1500 nucleotides, such as at least 1600 nucleotides, for example at least 1700 nucleotides, such as at least 1800 nucleotides, for example at least 1900 nucleotides, such as at least 2000 nucleotides, for example at least 2500 nucleotides, such as at least 3000 nucleotides, for example at least 3500 nucleotides, such as at least 5000 nucleotides, for example at least 10,000 nucleotides.

92. A method for screening for a compound capable of treating B-CLL, comprising administering a test-compound to a host cell comprising a recombinant expression construct, said expression construct comprising a constitutive promoter directing the expression of a polypeptide according to any of the claims 48 to 53 and on said cell measuring a parameter selected from the group consisting of: proliferation, apoptosis, necrosis, cell cycle changes or other physiological responses, inhibition of /activation of enzymes or caspases,

upregulation of/ degradation of mRNA or proteins involved in proliferation, apoptosis, necrosis or cell cycle changes.

93. The method of claim 92, wherein said host is a non-human mammal, such as a rodent such as mouse or rat.

94. A method for screening for a compound capable of treating B-CLL, comprising administering a test-compound to a cell line established from a subject diagnosed according to any of the claims 1 to 13, said method comprising measuring in said cell line proliferation, apoptosis, necrosis, cell cycle changes or other physiological responses, inhibition of /activation of enzymes or caspases, upregulation of/ degradation of mRNA or proteins involved in proliferation, apoptosis, necrosis or cell cycle changes.

95. A method for determining an increased or decreased predisposition for B-CLL comprising determining in a biological sample from a subject a germline alteration in a target nucleic acid sequence comprising 150,000 nucleotides, said target nucleic acid sequence comprising at least 10 nucleotides of SEQ ID No 1.

96. The method of claim 92, wherein said predisposition is a predisposition for poor prognosis of B-CLL.

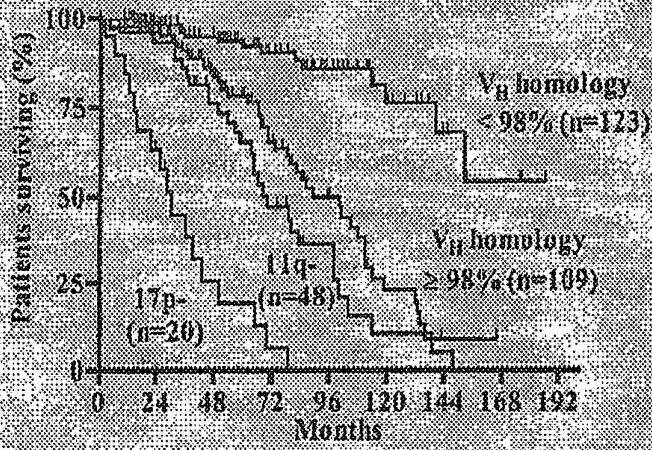


Figure 1.

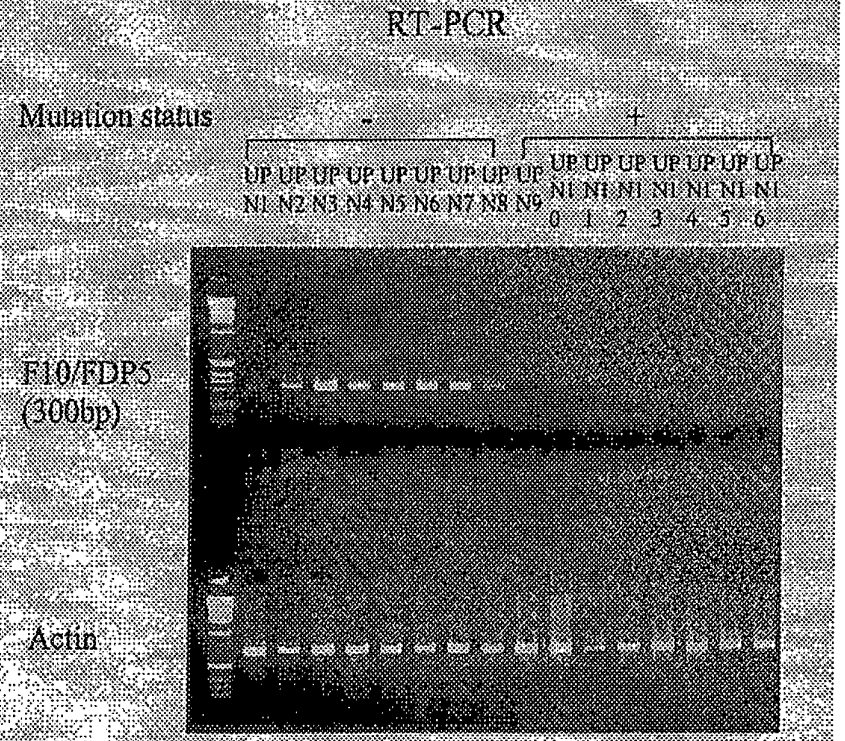


Figure 2.

Figure 3.

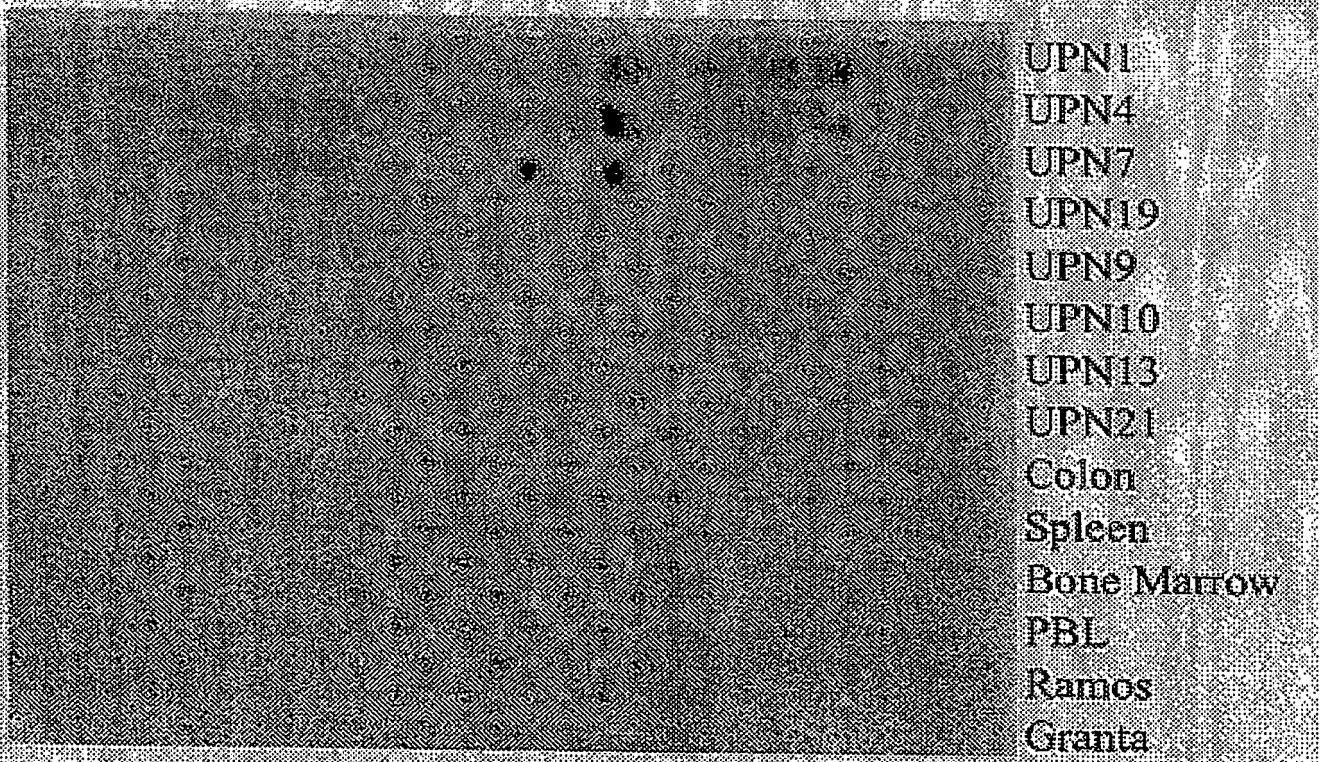


Fig. 4

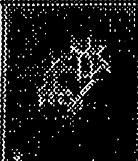
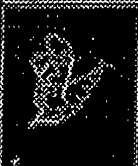
Fold library	Template Length	Model	[illegible] E_value	[illegible] E_value	[illegible]	[illegible]	Fold	SuperFamily	Family	Protein
d1nd212 23% i.d.	75		2.38	1	n/a	Alpha and beta proteins (a/b)	Thioredoxin fold	Thioredoxin-like	Glutathione S-transferases, N-terminal domain	Glutathione S-transferase
d1iara_ 17% i.d.	129		3.05	1	n/a	All alpha proteins	4-helical cytokines	4-helical cytokines	Short-chain cytokines	Interleukin-4 (IL-4)
c1dena_ 17% i.d.	89		5.34	1	n/a	not in SCOP 1.53	PDB header: hydrolase inhibitor.	Chain: A: PDB Moleculeiceberg (protease inhibitor).		PDBTitle: solution structure of in interleukin-3 1beta generation .
d1am9a_ 22% i.d.	80		5.57	1	n/a	All alpha proteins	Helix-loop-helix DNA-binding domain	Helix-loop-helix DNA-binding domain	Helix-loop-helix DNA-binding domain	SREBP-1a

Fig. 5

Fig. 6

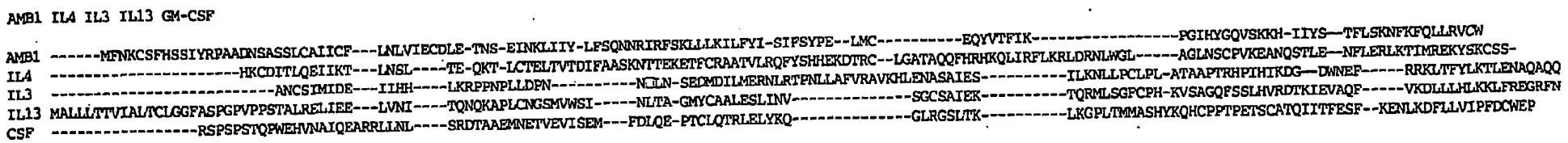


Fig. 7

Fig. 8

Sequence of ac063949.embum between 40,000 and 60,000. nrf1agalamb22.seq corresponds to (reverse): 58184 - 58408. CDS and exons are indicated.

```
40001  CACACGTAGG CTACGAGTGG CCCTCAGCCT GCCTCATCAT GGACCTGTGT
40051  TATAATAAAT ATGTTTAATT GTGCTGTTTT CTTATAGAGG AAAGTCCTGA
40101  TGTTAGTTGC CTTGAAGTCA GACACCCAGA GAGAATCACA GGTTTTCAGA
40151  TTAATTCATC GCTTGATTCT TATCCCTGAA GTCATATCTC TGGATCTCTG
40201  GTTCTCACAT TATAAATTTC AATGATTCTT TTTCTATATG GCCATGTCAT
40251  TCATATCCTG TGTAATATGG GGAAACTGAG GTATGAATGA CATCATTCAA
40301  AAAGCACCTG CAATTTTTCT TTGCCAAGCA CTTACAGCTT TTTCTCATGT
40351  TGCTTTCAAA AAGTCATTGA AATATTGTTC ACATATTTTG CAGATGAGGA
40401  AATGAATATT CAAATGCATT AGGTATCTTG TCCAAGTTCT TACAGCCAGA
40451  AAGTAGAGAA ATGAATTTGA ATTACAAATC TTCTACCTCT TGGCTTATGC
40501  TCTTTTCATG ACACTGGGAA TAAATGTCTG AACAAGCATG ACTTCATGTT
40551  TCAACTATTT ATCAAATACT TGTTTTCTAC TAAGATCTTG CACTCACTCA
40601  GTGGGATCCC CTGAAGCCTG CTGATTATTT GTCCTTTGGC ATTTATCACT
40651  CTCTGTGGGA CCTTACTCTC CTATGGTAAA GTTTTATTGT TATTAAAAGT
40701  ATTATTTGAC AATAAATGTA GAAATCCTAC AGATCATACT CAACAACATG
40751  TCTAATGTCA GCACACAATG TCTAACAATC ATTTATGAAT ACTTTATGTC
40801  AAACATAAGC AATAACCTAA TTAAGGAAGG TATTTTTAAT AAATTGACAC
40851  TTTTTGACAT AACCATATTT CAAGTGGCTC CATTGTTTTG TTTATTTATT
40901  TATTTATTTA TTTATTTATT TATTTTTGAG AAAGGGTCTC ACTCTGTTGC
40951  CCAGACTGGA GTGCAGTGGC AACATCATAG CTCACTACAG CCTCGACCTC
41001  TCTGGGCTCA AGCAATCCTC CCATCTCAGC TTCCCAAGTA GCTGGGACTA
41051  CAGGTGTGTA CCATCATGCC AGGCTAATTT TTCGTATTTT GTAGAGACGG
41101  GGTTTTGCCT GGTCGTCCGG GTTGGTCTCA AACTCCTGGG TGTTCCGCCC
41151  ACCTTGGCCT CCCAAAGTGC TGGGATTATA GGCATGAGCC TCAAGTGGCT
41201  ACTTTTTAGG GTTGAAATTT ATATTGACTG TCAACTAGCT TCCCTAGTTA
```

Fig. 8

41251 GTATTTGGGA TCTGCTAACT AATTTATATT ACCATCCAAC TTGTCAACAT

41301 TTGTTGAAAT ATAACTGTCC TCACTTTTTT TGTGTGAACA TTGAATACAC

41351 TTTCAGACTA AATTTGGTTT ATTACTTAAT GTCTTATTCT TTATTAGAGT

41401 TAATAATATT TCTTAATACT TTGCCTTCCA CAAATGAATA ACTTGTTTGT

41451 GATGGCTACC TCTTTTTTTC TCTTAGCCTG TCACAGGTAT TATGGATAAA

41501 AATTAGCACG GCTGGGCAAA AACAATGAAA GAAATACACT TGCCTGGGAA

41551 AGCTGGGGAG GGGTAAATGA ATATAATTCA AAATACCATA TATTTATTCA

41601 ACACTGTTGG AATATATGTC CTGTTGGAAA TGTAAAAGTG ACATATGTTC

41651 TCTTCCTGGG TCTCAGACTT TTAGGATCTA GTTGAGGGAA CTGGACTTAT

41701 ACACAAAATA CAATTCAACA ACATTATGAG CTAGAAAATC CATGAGCTAA

41751 AGTCTTTGGC AAAGACATTA GGTAACATGA GGAGTCAGGA AAAGGAGAAA

41801 TTACTGTGGG CTGGAATGGT CTGGGAACAT GAGATGGAGG AAGTGGCTTG

41851 TTACTGGAGA AAGGATGAGG TTCAAAGAGA TGGGAAAAAA AGAAAGAGAG

41901 AAGAAAGAAA AGAAATGAGG AAAAACAAGT TGCCAGAAAG AACAAGGAAG

41951 AATAGAGGCA GGTAAGCAGT GGATTTTGCC CTAGGGAAGG TAATATAACT

42001 AGAGACGGCA GTTTCTAACA GGCCATGATG AATAAGATAC ACTTTAGCCC

42051 TCATTGGTAC GTGCAGAAAT TCAAATTTGG AAATTCAAGC TTACATGACA

42101 GTAAATATAT GTTGGGAAAA AAATAACCGG TAAACATTTA CATCAGCTCT

42151 TTTTCCTAAA GAGAAACCTA TTCCATGCTA TGAAATATTT GTCACAATTC

42201 TGTTTTCAAA ATACTTGCTC TACTTTTCCA AGCCACAAGA GGAAACATTT

42251 TCTCTGCCAA CACTCTCTGA CCTTAACCAG TTTCTCCACT ACGTCTACTC

42301 TTAAGCTCTC TTTAGAGCTG TGTGTATCTC GTCTTTATGT AAACCTCCTA

42351 GATGATATAC TTATGGAAAT ATTCAGGCAA CTTTTTCATG AACTTTACCA

42401 GGAAAGACAT TTCTAGCAGG AGAGCATGAA TAGAAATGGA CTCTTCCCCA

42451 GTCTCTGCTG GGTTCTGACT GTGGTCACTC TAACTATAAA AAGTGTGTAA

42501 AAATCATGAG CAGATTATTT CATTTCCTTG GGGTCCCTAA AAATTTCAAG

42551 GTATCTGTAT TAGCACAGGA AGATTTAAAT TGATTTCTCA ACACATTCAG

Fig. 8


```
42601  ATATCTTATG AACTTTATTA AGATAAATTT CCTCCAGCAT TCAGAAACTC
42651  ATATATTACA GAATAAAAAA TAAAGCAGAA AATTAGTGTA CCTGGCTAAA
42701  AATGAGAGCA GGGTTCTATT TCATTTTGGA AAGTCACTAA GACAGTAATA
42751  ATACCATTAA TGATAAAATG TTAACATTAG TTAATTATTA GATGTGTTTT
42801  TGTATGCCAG CCACATAATA TATACTTTTA TATGTATCAC CTACATTTCT
42851  AGATGTAAAT GTGAGGGAAT TATAGTAGTA TCTACCTCGT ATGATTGCTG
42901  GATGATTTAA ATGAGCTGTT GTCTCAAAAA CTTGGTATAG AAAGCAGAAA
42951  CTTTTAGTTA TTAAGATTCT TACTATTCCA ATATTTGAAT AAAACAGTGA
43001  CCTGCTAAGA AACCCCAATA ATATTCTGAT ACATCAAAAC CTTCTGGCAT
43051  TAGATGTTTC TAATCTAACA TCTTCATATT AATTTTTTTA TGTTTTGATT
43101  ATCTACATTC AGTAGTGAAT GTGTTTCTAA ACGCTGGATG CATTTTTAAC
43151  TAAATGTGTT TTGTACCACA TTTTGACAAC TTTTGTTTTA ACTATGATTC
43201  AGCTTATAAC AAAACAAAAC AATGCATCTT CTCTCCACTG TTAATAAGGT
43251  TAATGAAAAG TTGACTTATG AAAAAAATCC TAATTTATGC ACATTCTCAT
43301  TGTTTTCCTT GCTAAGGATA TTAGTACTTG ACGATTCTGT AACAAAGAAT
43351  TATCATGGGA TGAAACTTTG ATGCAAATAT CTTATCAATA CAATGTGCTT
43401  GATTTTACCT AGATGAGATT TTTCTTTTCT TCTTTCTTTT TTGAGACAGG
43451  GTTTTGCTGT GTTACCCAGG CTAGCCTCAA ACCCCTGGCC CCTGGCCTCA
43501  AGTGATCCTC TCACCTCTGC CTCCCAAAGT GCTGGGTATT ACAGATGTGA
43551  GTCACTGAAT CCAGCCTCAC TTAGTTGGCT TTCTTAGTGA ATTATTTTAT
43601  CTGGTTCTAA AACTTTTTGA TAATACTCTC AAATATTTAT GGATTTTATA
43651  ACATAATTTA TGGATTACGT AGTTATGAAT TTCATAAATG ATTTTGTGAT
43701  ATTGCCACAG ATCATCACCA TTATACAGGA TGTATAACAT AACCATGGTT
43751  TAATATATTT TCATAAACTA TAGACCAAAC AAAGACTGGT CAGGACCAGG
43801  GCACGCATGC ATTTTATATG TGTGGTGCCT ATTGGAATAT GCCAGGCCTC
43851  CTGTGAAAAA AATCAGTAAG TGCTTATCTC ATAGGACCAA CGGCCCAACA
43901  TTCCTGAAGT CACTACCACA CTTTGCACTT ATCTCCATGT GGAAATAGAT
```

Fig. 8


```
43951  AGCCACTGTT GAATTCTGGT GAGAACGACA CGTCTGAAAT CTCTCAGCTT
44001  CACAACCCCT ATTACAGCCC TCAGAGAATC TTCTCACATA GCGCCAAACA
44051  ACAACTTTAG GAAGTGATGT TCCTAGAATG AATCAATTTC TAAAATTAAA
44101  AGTGAAAACA ATGACAAGGA GAAGGGAGGG TCAGAGAGGA AAGGCTGATG
44151  TTACTAAAAG ACAAAAGACA GTATAACCTC TTATGAGGAT GGTCCAGACA
44201  CTCAGGGAAA TGCAGGAAGA AATAAAAGAT AGGAGTTTGA ACCACACTGT
44251  GATGGCTAAC TTTATGTGTG GACCCGACCG ATCTATGGGA CACCCAGATA
44301  GCTTGTAAAG CACTATTTCT GGGTGCGTCA GTGAGGGTGT TTTTGGAAGA
44351  GATCAACACT TGAGTCAGTA GACTGAGTAA AGCAGATGGT CCTCACCAAT
44401  GTGGGTGCAC ATTGTTTAAT CTGTTGAGTG CCTGGATAGA CAAAAAAGGC
44451  AAAAGAAGGG TGAATTCCCT TTCTCGTCTT AAGCTGGGAC AGCCATCCTT
44501  TCCTACCCTC AGACATGAGA GGTTTGGATT CTTGGATCTT TGGTCCCAAG
44551  GGCTGACACT GGTGGCCACC TCTGGTTTCA GGTCTTTGGC CCCAGATTGT
44601  AAGTTACACC ATCAGCTTCC TTGGTTCTTG GGCCTTGAGA CTCAAGCTAA
44651  AATACACTAC CAGCTTCCCT TGTTCTCTAG TGTAGGGACA GCAAATCATG
44701  AAACCTTCTG CCTCCATAAT CATATAAGTC AATTCCCGTA ATAAATCTGT
44751  GCTTATATAT CTATAGCTTT CCTTTTGGTC TGTTTCTCCA AAGAACCTTA
44801  ATGTACACAC TATATGACCT AACCTGTAGT AATGATAACC TTATGCAGGT
44851  TTGAATAAGA TGATGGTATT CTCAGTATCT GGGAGGTATG GGCTAGAGTG
44901  ATGAACCACC GCCATGAGCC TAGGACTGAG GAGATTTCTG AAATGTGGAA
44951  TATTTGGTGT CAAAACCAAG AGATAATATA GCCATGTGGA AAACATGTAG
45001  AACTATCGTA TGATTCAGCA ACCCAACCAC TGGGAATTTA CCCAAAGGAA
45051  AGGAAACCAG TATATTAAAG AGAATCTGCA CTCCCATGGT TATTGCAGCA
45101  TTATTCTCAA TAGCCTAGAT ATGGACTCAA CCTAGGAGAT TAGATGAATG
45151  GACAAAGAAA ATGTGGCATA TGTACACCAT GAAATACTTA CCAGCTATAA
45201  AAAAGAATTA GCCAAAGCAG TGGTGTGTGC CTGTCATCCC AGCTCCTTGG
45251  GAGGCTGAGG TGGGAAGATC TCGAGGCCAG AAGTTTGAGA CCAGCCTGGG
```

Fig. 8

45301 CAAAATAATA AGACTCGGTC TCTAAAACAA TTTAAAAATA GGCCTTCCTT

45351 AAAAAAAGAA TAAAATCATG TCATTCACGG CAACATAGAT GGGACTGGAG

45401 GATATTACTG TTAAGTGAAA TTAGCCAGGA ACAGCAAGTT AAACCCCACA

45451 TATTCTGATT CATATGCGGA AGCTAAAAAA ACGTTGATCT CATAGAAGTA

45501 AAAAGTAGAA CAGAGGATGC TGGAGACTAG AAAAGGTAGG GAGAAGGAAG

45551 GGAGAGGGAA AAATTTGTTA ACAGGTACAA AAACAAAATT ACAGTTAGTT

45601 AGGGAGAATT AATTCCAGCA TCCTGTAGCA CTATAGGATG ACTATAGTTA

45651 ATAATAATAC TTTAATTAGT CTCAAATAGC TAGAAGGAGG ATATTGAATG

45701 TTCCCAACAC ACACAAAAAA ATGATAATGT ATGAGATGAT GGATATGGTA

45751 GTTATCCTGA TCTGATCACT CTACATTATA TGTATCAACA CATCACTATG

45801 TACCCCACAA ATATGTAGAA TTTTTATTTG TCCATTTAAA AAAGATAACA

45851 AATTTAAAAA TAAAATAAAA ACTAAATTAG TGTTCCATGT AAACCTGGAT

45901 GAACTGGTCA CCCTACGTCT GCCCATCTAG ATGGCTGGTC AAAGTTTCCC

45951 AGGCTCCACA TCAAGTTGTT CCACTGCTCA CTGGAACTTC CCTAGTCAGG

46001 TTGGGCAAAT AGTAATTTAC AGCAATAGTG AATTTATCAC TGACATTTCT

46051 TCAGTTCCCC TCTTTGGCAT CTGCTTCTTC TTTTCTGTAA TGCTGTTTGT

46101 TGAAATGCCC AACATTCTTT TTCTTCCCTA GAGCTATTCA GGGTGACCTT

46151 TCTTTTCGCA TTTTCCCATG CCACTTCCAT TATATCAAAA TAAAACAGTC

46201 CTGTGTGGCC ACTGCTCATG ACCTTGTTTC CTGCCATGTG AAGATAGGAT

46251 CGGCTGCTCT TTCTTCTCCT CCTTTTTTTT CAGAGACAGG ATCTCTCCCT

46301 GTCACCCAGA CTGGAGTGCA ATGGCACAGT CGTAGCTTGC TGCAGCCTCG

46351 AACTCCTGGA CCTCCTCAGC CTCCTGAGTA GCTGGGACTA CAGGTGCACA

46401 CCACCATGCC TTCCTAATCT GATATATATA TATATATATA TTATATATAT

46451 AAAATATATA TATAAATATA TATATTTTAT ATATTATATA TATAATATAT

46501 ATATTTTATA TATAAAATAT ATATATTATA TATATATATT ATATATAAAA

46551 TATATATATT ATATATATAT ATATATATTA TAGAGATGGG GTCTTGCTCT

46601 GTCACCCAGG CTGAAGATCA GCTGCTCTTT CTAATCTGTG GTTAGATAAG

Fig. 8


```
46651  ATCTGTCTCC CAGGGGATAA AATACTACCT GGAATAAAGG TATCTTTAAA
46701  ATAATCCCAG AGAAGAAAAC ATTTTTATAG TATGACAGAG GCAGAGAAAA
46751  CAGAGAATAT TTGTTAAGGC AGGACTTTCA CCACTCCCAG TACAATCATC
46801  TGTCTGTTAC CTGCATACCT TACACGGGCT GGCACTGCTG GGGGTACAAA
46851  GTAGATGCCA AACTTCACAA TGGTTAGATT CATGTTTAAA AAGCCATTGG
46901  ATCAAACCTT TGTGAAAGTT TCCAGCTTTT TTCTGTTCCA AATATGTGTC
46951  CATTATAAAA GAATCTCAAG AGCATAATTG CCAAGATAGT CTATGTCCAT
47001  GAGTATTTCA ACATCTCTCA TGAAATCTGT TCCCATCATT ACTCAAGATA
47051  TTGTATGAAC AGTATTCCAC ATAAACTAGG TGCTCAATAA TGATTGATTG
47101  GCCAATGGAG GGTCATTATT TAATGCACTA CAATCTTTTA TGCAAGGGGC
47151  CCACAGGAAT CAGTATGATC CCATAGGAAT CCTTTTCTTT TCCATTGAAA
47201  AAGAAACAGA TAGTGGCTTG TATTAGGTTT CTTGTGTGTG TTGTGAGGTG
47251  GAAAGATATG AAAAGAAATT TGATCAGAGC ATAAATCTGA GCCCATGGGA
47301  TAGGAAAGAA TGAGGGAATA AGGAAGAAAA CACAGATTAT AGACAGGAAA
47351  ATCAAACCTA TTAAAACTGA TAATTTTCGA ATACTAAAAA TGTACATTCA
47401  TTTGAACAAA AAGATTCTAT AAAGCAAGAT TTCTCTGTTC TTACCAGCAC
47451  TACCATGCCC AAACTACCTT AGGAAATGAA TAGCAGAGTC AAACTTAAAA
47501  GCACCTGAAA TTTAAAACAA AAACCAATTT ACATTTTATT TAAGAAAAGC
47551  AAACAGATGG GCCTGCTAAC AATGTCAAAG TCTCGTTTAC AAAGAAAAAA
47601  ACAAATCTGG AACCTGAAGT CAAACGAGTT CAAAATAAAA AGCAAACCAA
47651  TAAACAGAAA CCAACATAAA CAGAAGTTAC TACCATCTCC CTCAGCCTGT
47701  GAAATTCTGG AACTTCTCTT TCTTTCTCGC CTTCTTCTTC TCTCACCTGG
47751  AAGACGAGCA GAGTGAACAC ATCAGGGGTT GTCAGTTCCC CAGATGGCAC
47801  CACATTCATA AACCACCGAC TCCAGGAGAA TGTAGGAAGC TTAGTTAAGG
47851  CCAAAGTTCT CTTTGGATCT TCCTCATGGG CTTCAAGGCA AAAGAAAAAA
47901  AAGTTTGCTT GAGAATATCT TCATATCTAT TAGTTTGAAC CATGCAAAAT
47951  TACAGTTTTT ATAGGTAAAA TGAGTGCATA TTGGCAATTT CAAATGATTA
```

Fig. 8

48001 ACCCTAATAC ATTATGCTTT TGGGTATAGA AATATTCAGA TCTTAAACAT

48051 ATGCTGTTAC ATACAAAATC AGGTATATTC CTGCTTCTAT AATTAAAGCA

48101 AAGAGAATTT CTTTTGGTCA CTACTCCTTC TGACATGAGG TATGAACCAA

48151 GTTCAGGACC CCTAAAGGTC TGGGTCTGGG TCATTTCTCC ACCTCTAACT

48201 TGTGCCGCTT TCTTGGTCAG TCATTGTGTT CTGAGCTGTC TCATAAAACA

48251 TCTGCTATGA CTTTACTTTC TCCTGATAGG GTGGCTTTCC ATCGTTGGCA

48301 CTTCGTTGGC CTTATTGGTA TGCTTTATAC ACTGGTTCTC GTTTCCAAAT

48351 TGGCATTATT ATTGTTATGA TTCCTGCTGC TCTCCCACAT TTCCCATCTT

48401 TCTCCTGATC TCTCTCACCT GTACATTTCT TACATTTTCT CCTGTGCTTC

48451 CTTCTTCCCA TCATCATTGC CCAAGTGTGT CTTCTTTCTT CTCCTTGTCA

48501 CATTTCCTTT GCCCGCTCTC ACATATGCAG AGATGGCTCT TGGTTTTCCT

48551 TCTGAAATCT CATAGTTTGG AGGTAAACTT GTTAGCAAGG CCACTGAGAA

48601 GAGAACAAAA GGGAAACATA AGAGAAACCA AGTCACTATC TCTCTCATTT

48651 CCTGGTTTCT AGAAGTAAGA CCCAAAGAAC TCACTGTTTC AGTGCTTTCA

48701 GCTCAGGCCA AACTAGGGTG ATCAAACTGA GCTTCTGAGT GCTGATCAAA

48751 ACCTATAAAA CCAAGTAGAC AGACCATCTA CAAATCTTCA CTGTTAAATA

48801 CCATAAAGAA TGAAAAGGTC ACTAATTGGT AAGACTATAT GTGTGATAAT

48851 TAAATTTATG CATCAACCTG GCTAGGCTAA AGGATGACCA GGTAGCTGGT

48901 AAAACATTAT TCTGGGTGTG TCCATAAGAG TGTTTTCGGA AGAGATCAGC

48951 ATTTGAATTG GTGAACTTAG TAAAGCAGAC GGCTCTCACC AATAAGGGCA

49001 GGCATCATCC AATCTGTCGA AAGCTTGAAT AAAACAAAAA GAGGAAGGGA

49051 AAATTTGCTT CTTTTCTTCT TGATCTAGTA TATCATCTTC TCCTGCCCTT

49101 **GGATGTGAGT GGGCCTTCAG ACTTAAACCA GGAGTTACAC CTTTGGCTTC**

49151 **CCTGGTTCTC AGTTCTTTGG ACTTGGACTG AATTACACTG CCAGGTTTCC**

49201 **TGGTTCTCCA GCTTGCAGAT GGCAGATCAT GGGACTTCTT GGCCTCCATA**

49251 **ATTGTGTGAG TCAATTTCCA TTTTATTTAC ATATCCAGTT ATGCATTGCT**

49301 **TAACAATGGA GACAGGTTCT GAGAAATGCA TTGTTAA*GTG ATTTCATCAT***

Fig. 8


```
49351  TGTGCAAACA TCATAGAGTG TAACTACACA AACCTGGACA GCATAGACTA
49401  CTACACATCT AGGCTACATG GTGTAGCTTG TAACCTCATG ATAAGTATGT
49451  ATAACATCAT GATAAGTATG TATGTATCTA CCATATCTAA ATGTAGAAAA
49501  GGTACAGTAA AAATATGGTA TAATCTTATG GGATCACCAT CATATATGCA
49551  ATCCTTTGTA GACTGAAATG TCATTGTGTA GTGCATGACT GTATACGCAC
49601  ACATACACAA ACACACACAA ATATACTATT GGTTCTTTTT CTCTGAAGAG
49651  CCCTAATACA ATATGTTATA CATTTATATT GACTCTATTT CAAAATTTAT
49701  GGTTTTGGTG AAACATATGT GGAGATGGGG CATAGGTGTG TGAACTGGGA
49751  TAGTGTCCTG CTGATGAATG GGTGGGAGGC ATCATTTGGG ACAAGCCCAG
49801  GGCATCAGCT TATAGATATC AAGAGCTCAA CAAGAGCACT TTATGGCAAA
49851  ACCTCCCACA AGACCTCTCA GAAGTTGAGA AACTGCTAAA AGTTTCTTTA
49901  TGACAGATGA CATTTATGGA TAAAATAGGG ATTAGCAGGA TTCTTTAAAT
49951  ACTTTCGAAC ACTAACCTTC ATTTCTACCA GGCAGTGGGG CCCCAAGTGC
50001  AGGGCCATAG GAAGTACAAG TCTGGGAGAT ACTAGGCTGC ACTGTCTGTA
50051  GAGAATCTGA AAAAATAATA GAGTCACTGA AATGCAGTTT GGTATAATTA
50101  TTGCCATGCA TCATAATTCT AAATCATACT AGTGGTCAAA TACTCTTCCC
50151  TGAAAAAACA TTTTCTTGGT TTGAATTCTA AATAATTGTT GTGGTCACCA
50201  CTGAGCTTTT AAATATATAA ATACTTTCAA GTTTGCATAT TTTTATTACC
50251  TGTTCCTTAA CAAACATTGA ATTCAACATG AAAATGATTA TGGGAAACAT
50301  TCGGGTATAC AGTCCCTGAC TCTTAAGGAC TCAGGTAAAT ACTTAGGGTA
50351  TTTCATGGCC CTAGTCTTTG GGGTACCACA TGTTTCTTCT TCAAATCACA
50401  GATTCAAAAT CAAGAATGAT AACACAGTGA TTGTGTAGAC AAAATAAGTG
50451  AACCAAAATT GCTTGCTTCT GTCATTCTAT GGAACCACTG AGAGTTTTTA
50501  CTTGTGCTTA AAATTTTGAA TAGTAAAACA GAGTGTCAAC TTCATGCTGG
50551  AATATTTTTG GCTTTTTAGA CACAATTTTA AGTACATGAA GTATTTTTAC
50601  AAGACTAAGT AACATCACTG AAATTACAGC TTTCTTCTTT TTAAAACTGG
50651  TATTTGTTAT AAAACTAAAG AGCGAATCAA GAAAAGCATA ATTATTACTG
```

Fig. 8


```
50701  ATTATTACAG GATTATTACT GAAAAAGAAA TGTACGGAAT AGAGGAGGAA
50751  GGAGTTAACA AATGATCCAC TCTGGGTGTT GAAAACACCA ATAAGCCTGC
50801  TTCCAGGAAG TGCCTAAGAC AGAGCTGGCT CAGCTTGCTG GGTCACAGCA
50851  TGTAAGGAAA CTGCTGGGCT ACATGCCACC ATCCTCAGTT GTCCAGATAG
50901  ATAATCCCAT AGCCCCATGG GGAAATAATC TTTAATTATG ATATAGCTGA
50951  CACCATTCAA AGCACTATGC TAAGTCCTTT ATGTGAATTA ACTTTTGTCA
51001  AATTTATTTT TCATAAATAA CCCAAATATG TATACCACTA TTATCCTACC
51051  TTAAAGAGGA GAAACTGAGC TCCTAAAGTT TAAATATCTA ACCCAAGTTA
51101  AGACTGCTAG TCACCCTAGG CTATTAACTC AGGCAGTCTA ACTCAGGTAT
51151  AATAACATTA TGCTACTGTT TGCAGCTTTG ACTATGCCTG AATTATAACG
51201  TCATGCTATC TAACTAAAAA GCTAAGGGAA ATAAAATGAG CCATAGGGCT
51251  CAATTTCATA AAAGGAGAGA AAATACTGGG GAAAAGTGAT AATGCAGAGT
51301  TTAAAATATT TTTGTAAAAG TGCCAGAGAT TGAGTATAAC AAGTGTGACC
51351  AAAAAAAAAA AAAAAAAAAA AAAAGGAAGA AGGTAAAAAA AAGAGGGAGG
                              RACE end
51401  TCTGAGAAAT AGAAATATCA GAGGAAGGAA ATAAAGGAGG GTGAGAGTAA
51451  ATTCTCTTTT AGCATTCAGA TTCCACAGAT TCCACAAATC ACATTTCTTT
51501  TTTTACCAAC TAAGGAAAAA TAACACTTGA CCTAACATTT CATTGCAGTT
51551  AGCTAAAGGA TGCTAGAAAA ACTATGTTGC AGTGGTTTGC TCTAATTTCT
51601  TCAGGAATAG AGAAAAGTGA CAAAAAGATC AGAGAAGAGA AGAAAGGAAA
51651  CTATCAGAAA AATACAGAAT TGGAGTAGGA TATAACATAT TTGGGTTGAA
51701  GGTAAAATTT TATATTGTAA TCTTAAGTAT CTTGCTACTT CAGTTTGGTC
51751  CCTGGAACAG CAGCATCAGA ATCTGCCGAG GGCTTGTTAA AAAGGCAGAA
51801  TCTCAGGTCC CATCCCAGAC TCACTGAATC AGAATATAAA TACTGACAAG
51851  ATGCCCCGGG ATTCATATGC ACAGTAGAGC TGGCGAAGTT CCATTGTAGC
51901  CTGTGATTGT TTTCTGCAAC TTAGTATTTC TGAGTTTTCC CAAGGAAGAA
51951  AACCCAGGCC TTAGCTTCTG GCAGACTTGT GTTTCTCCTT TACTTACTAG
52001  CTGCATGACT CATGAGCAAG GAAATCAAAC TTTATGTGCC TGAGTTTCCT
```

Fig. 8


```
52051  CATCTATAAA ATGGAGACTA TAATAATCAT CTCCTAGGCT TGTTTTGAGG
         M  F  N  K     C  S  F     H  S  S     I  Y  R  P     A  A  D
52101  ATGTTCAACA AATGCTCCTT TCATTCCTCT ATTTACAGAC CTGCCGCAGA
         N  S  A     S  S  L  C     A  I  I     C  F  L     N  L  V  I
52151  CAATTCTGCT AGCAGCCTTT GTGCTATTAT CTGTTTTCTA AACTTAGTAA
         E  C  D     L  E  T     N  S  E  I     N  K  L     I  I  Y
52201  TTGAGTGTGA TCTGGAGACT AACTCTGAAA TAAATAAGCT GATTATTTAT
         L  F  S  Q     N  N  R     I  R  F     S  K  L  L     L  K  I
52251  TTATTTTCTC AAAACAACAG AATACGATTT AGCAAATTAC TTCTTAAGAT
         L  F  Y     I  S  I  F     S  Y  P     E  L  M     C  E  Q  Y
52301  ATTATTTTAC ATTTCTATAT TCTCCTACCC TGAGTTGATG TGTGAGCAAT
         V  T  F     I  K  P     G  I  H  Y     G  Q  V     S  K  K
52351  ATGTCACTTT CATAAAGCCA GGTATACATT ATGGACAGGT AAGTAAAAAA
         H  I  I  Y     S  T  F     L  S  K     N  F  K  F     Q  L  L
52401  CATATTATTT ATTCTACGTT TTTGTCCAAA AATTTTAAAT TTCAACTGTT
         R  V  C     W  *
52451  GCGCGTGTGT TGGTAATGTA AAACAAACTC AGTACAGTAG TATTCAGTAC

52501  AGTATTTAAG CCCCTGTACT TAAACATATT CCTCGTACCA ATGAAGTTAC

52551  ATGAAAAGCA AATTTGTGTG AGATATCGTA GATGGAAGTA AATTAGTCTT

52601  TATGTTCCCC ACAAATTGAA ATGCATTTCA AAAACTCTGT GTGTGTATGT

52651  GTGTGTGTGA CAGAGTGTGT GTGAGAGAGA GACAGAGAGA TACGCTTTGG

52701  TTGCCTCCAT AAGCTGGCTG CTATGATTAA TAAGACCAAG TTTTCTAAAG

52751  AAAATGAGAT CATAACAAAA GCCCTCTTTA TGACTATCTT TTATCAGGGG

52801  CAAAAAGGAA AGAGACAAAA CAGCATGAAA TGATGAGACC AAGTGATGAA

52851  AATTCATTCA CAATGATTGC TTTCAAGAGT AATTTCTCTT GGGTAATTCA

52901  GCAGCCTGTT ACTATGGCTC TCTGGAGTGA TAGCTAATGT AAATGAAGCC

52951  TCTAAAAGTG GATTATCCTG ACAAGAATAT ACTCAGCCAA TAATGCAACA

53001  GAAATCCATT CAAAGCATTC GGGAAAAATT CAAAAGAATA AATATTCTTT

53051  TTTTTTTTTT AAAGTTAATG ACCTACGATC CATTTCTTCC CTGACTAACA

53101  AGCAGCAAGC ACTTAAAAAT ATCCAGCCAG GATGAAATAG AAACCCACCT

53151  GACTTGTTAA TATTTTTGTT TGGTCCCAGG GACTCAGATT CTAAGCCAAA  Exon

53201  TTCTTTGAAT GATCTTGGCA AATGTCTCGA ATTATTTTTG CCAACTTTTC

53251  TTTATCTTGG AAAAAAAGTT TCATGAATGG GTGTCAAAAT TGATTAGTTT

53301  TAAAAACCTT TCTTGCAGAT ACGTATGGCA CCCTAAAACT GTATTAGAAA

53351  AAAAGTAAGT ACTCTGTAGT GTGAAAAATT CTTAAAGGAC ACCCTCTTTT
```

Fig. 8


```
53401  ACAAACTCAC AAAAACAGCC TTTGGAATAC CCACATGAAG TAGCTGTTGT
53451  TATTGCTTTC TATATACCTA CATCTTGTCT ATTATAAAAA GACTGGTTTT
53501  TGGCAGGTGT GGTGGCTCAC ACCTGTAATT CCAGCACTTT GGGAGGCCAA
53551  GGCGGGCGGA TCACCTGAGA TCAGGAGTTC AGGACCAGCC TGATCAATAT
53601  GGTGAAACCC AGTCTTTACT GAAAATACAA AAATCACCCG GGTGTGGTGA
53651  CGGGCGCCTG TAGTCCCAGC TACTCGGGTA GCTGAGGCAG GAGAATCACT
53701  TGAACTCAGG AGTCAGAGGT TGCAGTGAGC TGAGATCATG CCACTGCACT
53751  CCAGCCTGGG TGACAGAGCA AGACTCCATC TCAAAAAAAA AAAAAAAAAA
53801  GACTGGTTTT TCAACAGCTA TTCCCACCCC TCTGCATGGA AATATTCACC
53851  CAGTCAATTG TTTTCCTAGT TTGGGTAATG GCCCTCTGGG CAGGACTGGA
53901  GTGGGGCACA CAGGAGAAGC TGCAAACTAT GTTTAGAAGC ATGTCTGGGA
53951  AATGTCATGC AAGAAAAGAC ATATTTAAAG GTAGGCTTTG CATGAATGGA
54001  AAAGGAGAGT AATTCTATGT AGAGCAGAGC CTCTTACTTG CAGTGAGAGA
54051  AGCAAAAGTG GGGAAGCAAG AGGAATTATG CTTTTCATCA GCCAAATTTG
54101  CAGGTAGGAG GATTGGCTCA GTCATCTTGG CTGAGGCTCA TGAAACCAGG
54151  TGTAAAGAAA GTGGACTAGA TTAATTTCAT CCATTACAGG AAGAGGAGCC
54201  GTGAAAGATA ATCCAGAAAT CATTGGGATT TGATGGTAGA AGGTATTTTG
54251  GGACTATTCC ATTTGAAATG AGAAGGTACC TGACATTCTT TGAATTCCTT
54301  TCAAGCAAAG GATTAAATTT ACCCATGAGT TGACTCAGAA AAAACATAAA
54351  AAGTATTGTT GCTCTGCTCA GAGTTTTATC TAACTCATTC TCACTTCTTA
54401  TTCCATGATG AAATGACATA AATGAGGTTT TTTATTGTTG TTGTTGTTGT
54451  TTTCTGGACA CAAGGCAAGG TAGCTACCTG GGCAGAGCTG TTTTATTTCT
54501  CTATGCCGTG GAGAGAAATT GGTTAATTGG CCATGGAAGG CAGTCATTAA
54551  GATGTTCCCA TGCGAGTGAA CTTTCCAGGG TTCCCAGCTT CTGCATCCTT
54601  CCCTGTCCCT CAATTCCATT GTTGGTGATG ACAATGTCTC TCCCATCAGC
54651  CTCATGAAGT TCTCTCTCAT TTATTAAAAT TTGCTTTCAG GAAAAATTTT
54701  GAAAATGTGT CCAGTAATGC CTGATTGGCC CCTTATCCTA AAGGCTTAAA
```

Fig. 8

54751 CTGGAGGAAG GAAGCTAAAC TGAGAAATCT TGCAAATCAT TGAGCCAAAA

54801 ACGTATTAAT AGCAAGATCT ATCATTTATT GACTAGTATG TGGCAGGCAG

54851 TGCCCTTTTA TTTAGGCAGG GAGAGTTGAT GGGGGGGGCG GGGTTCACAC

54901 ATCTTAAAGA GGTGCTATCT CCTCCTATAT AAATCATGTA AGTCAAGAGA

54951 GTAAGGAATT GTCTTTGTTT GGTTATATTC AGGGGATTAG AGTATACAGT

55001 AGAAGATCCC AAGAAACCTT GGGATCATTT TAGACTAAGA AATGCCAATA

55051 CCGCCGGGCG CGGTGGCTCA CGCCTGTAAT CCCAGCACTT TGAGAGGCCG

55101 AGGTGGGCGG ATCACAAGGT CAGGAGATTG AGACCGTCCT GGCTAACGTG

55151 GTGAAACCCT GTCTCTACTA AAAATACAAA AAATTAGCCG GGCGTGGTGG

55201 CGGGCGCCTG TAGTCCCAGC TACTCGGGAG GCGGAGGCAG GAGAATGGTG

55251 TGAACTCAGG AGGCGGAGCT TGCAGTCAGC CGAGATTGCC CCAATGCACT

55301 CCAGCCTGGG CGACAGAACG AGACTCCGTC TCAGAACAAA ACAAAAGGAA

55351 ATGCCAATAC CAGCAGAAAT AGAGCCAAAT CATGAACATA AGCTAAACAA

55401 ATGTTGGCAG TGTAGCCTAG TGGTTAAGAG AGCAGACTCT TAACTAGAAC

55451 ACTGCACTCC ATGTCCTCAC TGTAGACCCT CACTGTGGGG TTCTAATTAA

55501 CCCCTGTTAC TTACCAGTGG CAGTCTTAAG GCATTCCTTA AGTTCGTTGT

55551 GCCCCAATTT GTTCATCTGT AGAAGGGGTA GGATGACAGT AGTGTTTACT

55601 TTATAGGCTT ACTGTGAGCA TTAAATGAGT TACTACTGTA TTTGTAAAGT

55651 GCTTAAAATG CTGCTCCAAA AGAGTTTGTT AAACACTTAA GAACTGATTT

55701 ACTTGCATCT AAACTGACAG CTCTCAATAA CTGGAAATGA TCAAGCATAG

55751 GCCCTGGAAT ATAAGCAGGT CTACATGAAG GCAAAAATGT TCGTTTCTTT

55801 TGTTCAGCCC TGTGCCTAGA TCAATATCTA GTGATCATGC TCAAGAAATA

55851 TTGTTGAATG AATCAATGAA CCTACCGAGG TAGTTACATA AAAGAGTTCT

55901 GCATGAGTAC AAATCTGGGC AAAGTGACCT CCAAGGAAAT TTCCACTTTT

55951 AGATTCTGTG ATTTCCTTAA GGAACTGATA AATTGGTGTG ATACAATGTA

56001 AAAAAATGTG CCTATATGAT TTGAGAAAAA CTTATTTTCT CTCCCTCTTT

56051 TTTCCTTCCT TCCTTCCTCC CTCCCTTCCT TCCTTCCTCC CTCCCTTCCT

Fig. 8

56101 TCCTCCCTCC CTCCCTTCCT TCCTTCTCTT TCTTCTTTTC TTTCTTTCTT

56151 TCTTTCTTTC TTTCTTTCTT TCTTTCTTTC TTTCTTTCTT CTCTTCTCTT

56201 TCCTTTCTTT CTTCCTTTCT TTGTGCCTTT CTTTCTTTCT TTCTTTCTTC

56251 TCTGTCCTTT CTTTCTTCCT TTCTTTCTTT CTGCCTTTCT TTCTCTTTGT

56301 TTTTCTTTCT TTCCTTCTTT TTTCCTTTAA GCAGACCATG TCTGTTAGAT AMB-UP-5

56351 GAATGCCTTT TTCTAGTTAA AAGGTTAAAC AGGAAAGTGA AGCACAATTA

56401 TCAAGGGTCT CCAGTCATCT CCACATGTTC TTAATCATTA TCTTCTTTTA

56451 CAGTTTCATA TCTCCAGGCC TTTCATTGGG TCAGGTTGGC ATTTCGCTGC Exon

56501 CCTTTATGTG TGTGACAAGT GAAAATAAGG AAAGAAAAAA ACTCAAGTGA

56551 AGAAAATCAG AATCTGCGCA GCAGTTCCTG GGCGTTTCAG CTGCTTCCCA

56601 CATCACCTGC CTCATCAAGC CCCAGCATCC ATCTCCTTGC TCATCTTACA

56651 CCCTGTGTGC ATGACAGGCC CACCATTCAT TTATCAGAGC AAAGGCTCTC

56701 CCACTATTCT GGTTCACCCC CCTACTTAGC CAGATATACA AGAATATCTG

56751 CACGGATGAC CTGCCTCACC TGGGAGCTCA GAGGAGCTCA GATTCCATTA AMB-UP-4

56801 CTATCGCACC AAGGACAGAT CTCCCAGCAA GAATGACAGA AAAGACTAAC

56851 TGCCCCCAAA ATCTCCCTTC CAAAACACAG TTCTCTTAAT TCTCCCAAGA

56901 AACCAGAATG TGACTGCTCA CCTCTCTAAG GACCTGAAAA CAACTGGCCA

56951 TTTCAGCTAT TTAAATCAAC TTTAAAAAAT CCAACCGCCA AAATATTAAA

57001 CCATTTTGGT TGGAATGATA ACATAACTAA CCTGCTGACA GCTGCTTCTG AMB-DP-4

57051 CTAGGTGCAA AAATGGAAAA AAAAATACTT CTAATCAGGT CAAATCACTC

57101 TACCTTTGGG ATTCTAAATT TACTCATATT CTCAAAGAAA TATATTCAGT

57151 CATAGTGGGG AAAATAGGAT TATTCCTTTA GCTCGATAAG CAACCAGAAG

57201 TTCTTCCTTC AAATCTTGAC ATTTAATCAA TCAGAAATTG ATTTTTGGAA

57251 AACTGTTTCC TATGAAGCTA TCTCTGCCTG AAGGATTTTT CTTTTACAAT

57301 CCAGACTATA GAAGGAAATT CACAACCTGG ACTTTCACCT CCATTGGTCA

57351 GAGTTTTACT GACCAATTCC CACCTCTGCC TTACACCTAA CGGAAGTTTA

57401 TGCCTGTTTT CTCTTCACAT ACCCCAACAG TTACAAATGG TTGTTATTAT

Fig. 8

57451 TAAGCATCTT TTATTTTGTG GCCTCTGATT ACATGGTCCC CTAAATTTTG

57501 ACCTAATCAC AAAAGATTGG TAAAATTTCT TAACATATTA ATAATATTTT

57551 GTTTATGTGT CAATATCTTA GCATGTATCA ATTAAGACAG AGGTCTTAAC AMB-UP-3

57601 GTTCTCTTTT TGAAAGAGAA TATTAGGATT CAGAGATATT AAGAGATTCT AMB-DP-3

57651 CCCAGGATCA CAGTTAGGTA ACAGAGCTGG ATTTTAGTCC AGGTCTGTCT AMB-cDNA-1
(Rev + Xho1 site)

57701 ACAGCTCTAA CGTATATACA CCCTTTGTAT AACATGTCAC GAATTCAGCA
G G

57751 TAAAGGGATC TTCAGTGATC TAAGTCAGGG GTCAGCAACC TTTTCTAAAA

57801 AGGACCAAAT AGTAATATTT CAGGCTTTGT GGACCCTATG GTCTCTATCA

57851 TAACTGTTCA AATCACCATG TAGTGTAAAA GGAGCCATAA GCAAAATATA AMB-UP-2

57901 AACTAACGAA TGTGGCTGTT TTATGGGATT TTTTTTTAAC TCTTTATTTA

57951 CAAAAGCAGG TGGCAGATCA GAACTCACTT ATGGGCCATA GTTCTCTGAC

58001 CCCTGACCTG AGAAAATCTT ATATTTATGG ACAACATTTA GACTGTGACT

58051 TGCCAAGTAA GAACAAGAAG CTCTGTCAAC TGAAGGTCAA GGCTGGAGTT
+T3 AMB-T3

58101 CTGAAAGCAA AGAGCTGTCT GGTGTTAATG ATAAGTGAAA TAGTTAAAGT

58151 TAGAAGATCC CAGTTATAAG AAGCACAAAG AATAATGACC ATAGACTCCT AMB-UP-1
AMB2-T3

58201 GAACAAGAAT GTCTGGACTT CTGGCTTAGG CACTCTTGTT GTATGGTCCA

58251 GGCCAAGTTA CCTAATCTCT CCAGGCCTCC ATTTTCTTAT CATTAAATGA

58301 AGATAATAAA AGTATTTTCC TCAGAGAGCT GTAAGAATAA ACTGAGCTAA

58351 CCCATGTCAA GCACATAGAA TAGGGCCCAG CCTATATTAA TTTATCAATA AMB-DP-1&2 Rev

58401 AATGCCAG-Poly-A
(CCTATTCTATGTGCTTGACATG AMB-DP-2) inside 3'-end
(ATTGATAAATTAATATAGGCTGGGC AMB-DP-1) 3'-end

CT ACATATTAGT TCTCTATATT TTTATTCATT ATCATAAAAT

58451 GTTTATCTAC AGATTGGCAT TGTAAGGATG GAGTTAAAAT TGTATGTATG

58501 TGAAGGGAAA TTATTCCTGT TACTATTGAT CTGCATCACA TTACCCCAAA

58551 TTTGATGGCT TAAAGTAACA ACATTCATTT TGCAAACAAA TTTGAAATTT

58601 GAGGAGGGCT TGTCTGGGAA GACTTGTCTC TGCCCTATGT GGTATCAGCA

58651 GGGGGAGGCT TGACGGACTG GCACATGCCC TTCCAGAATG GCCCACTCGC

Fig. 8


```
58701  ATGCCTGCCA AGTTGGTGCT GGCTCTTGGC TGGGAGCTCA GCTGGGGCTG
58751  AGTGCTAGGG TCCCTGGGAG GTTCCTTGTG GCCTGAACTT CCTCACCACA
58801  AGGCGGCTGC GGTGCGAGAG TGAGCATTTC AAGATAGAGC CAAGATGACA
58851  CTGTATTACT GTGTAAGACC CAGCCTGGGA ATTAATGTAG CCTCACTTCC
58901  ATCCCACTCT ATTTTTAAAA AGTGAATTAT TAAGGTCACC CCATATTCAA
58951  GGGGATAGGA ATTAGACTTC ATCTGTATTA AGAAAAATGT TTTTAAAAAT
59001  TGTAGACATG TTTTAAAATT CTAAAGTCCA CTTACTGGCT GCAGATTATT
59051  TATATATACA TGCAAGATAC ACTCCTACAT TCTCTTCTTA GAAGGCTCAG
59101  TTGCAGGTAC AGATGAAGCT CTTCAAGTGA GATTTCTTAT GTATTTATCC
59151  TCTCAATCTG AAGACTTGTA AACTAAGAGA CAAGTTATTT GCAACCTACA
59201  TACGCAATAT TCAATGGTAA AGTATACATA GGACAGCCAC TACAGACACT
59251  CTTGTTTTAA ATAGAGGAAA ATGAGAGCAC ATAACAGTCA TTGGCTCATA
59301  GCAACTCTGA TATCCAGACA GCAAACACAA GCAGGTCTTT TTTTAGGTCT
59351  CAGTCCTACT GCCTGGATTC CCTACTGCTC TTGGGTCTTC CCTCCAGGTT
59401  CTTGGTTCTT GGACCTCTTT TCATTTAATA CTATTTCTGT TCCTTTAAGT
59451  TCAAGCTGGC AAAATATGAT TGTACAATTC TGTTTAAAAT TCCAGGACTT
59501  CCTGTGATTC TTATTGGGGA ATACTCCATT AGACAAGAAT CTCTTTGACA
59551  TAAGCCATTC TCTACCTGAG ATCCCTGTAA GGCTGTGATG GGACCACATA
59601  ACCTTAAAAT TATTAGAAGA CTCATTGTTT ACTGAGAGAA TATGCCTAGC
59651  ATATGCTTAG ATCCTTAGAG GAACTCTGTT TCAAAGGGCT TATGAGACAT
59701  TACCTTATAT CTTTCTAAGG TACAAACAAA AGGTCTTTGG CTTTTGAGTT
59751  TGATCTTTGA GCTGACACCT TTTCTTAATT TGAGAATCCC CTGCTCTATG
59801  GAGAGACTGA CAAAGAGAAA TAGTTTTATA TTTGAATGTA ACATCTTGGA
59851  TCTTTAATAG ATTATCTTAA AATTTTCCTG AAAATGTAAC AGTTCCTTTT
59901  TTTAAAATTC ATTCTCCCTA CACACTTATT ATATATGACT AAAAGAAACT
```

AGTCCAGGTCTGTCTACAGCTCgAgCGTAT

ATACGCTCGAGCTGTAGACAGACCTGGACT

Fig. 9

AMB1 mRNA Longest form (SEQ ID No 4). Short form (SEQ ID No 2) starts around pos. 2317
Coding region: 3001 - 3363 Stop codon 3364-3366
Position of intron 4254. Intron length 3099 (not included)

```
   1  GGATGTGAGT GGGCCTTCAG ACTTAAACCA GGAGTTACAC CTTTGGCTTC

  51  CCTGGTTCTC AGTTCTTTGG ACTTGGACTG AATTACACTG CCAGGTTTCC

 101  TGGTTCTCCA GCTTGCAGAT GGCAGATCAT GGGACTTCTT GGCCTCCATA

 151  ATTGTGTGAG TCAATTTCCA TTTTATTTAC ATATCCAGTT ATGCATTGCT

 201  TAACAATGGA GACAGGTTCT GAGAAATGCA TTGTTAAGTG ATTTCATCAT

 251  TGTGCAAACA TCATAGAGTG TAACTACACA AACCTGGACA GCATAGACTA

 301  CTACACATCT AGGCTACATG GTGTAGCTTG TAACCTCATG ATAAGTATGT

 351  ATAACATCAT GATAAGTATG TATGTATCTA CCATATCTAA ATGTAGAAAA

 401  GGTACAGTAA AAATATGGTA TAATCTTATG GGATCACCAT CATATATGCA

 451  ATCCTTTGTA GACTGAAATG TCATTGTGTA GTGCATGACT GTATACGCAC

 501  ACATACACAA ACACACACAA ATATACTATT GGTTCTTTTT CTCTGAAGAG

 551  CCCTAATACA ATATGTTATA CATTTATATT GACTCTATTT CAAAATTTAT

 601  GGTTTTGGTG AAACATATGT GGAGATGGGG CATAGGTGTG TGAACTGGGA

 651  TAGTGTCCTG CTGATGAATG GGTGGGAGGC ATCATTTGGG ACAAGCCCAG

 701  GGCATCAGCT TATAGATATC AAGAGCTCAA CAAGAGCACT TTATGGCAAA

 751  ACCTCCCACA AGACCTCTCA GAAGTTGAGA AACTGCTAAA AGTTTCTTTA

 801  TGACAGATGA CATTTATGGA TAAAATAGGG ATTAGCAGGA TTCTTTAAAT

 851  ACTTTCGAAC ACTAACCTTC ATTTCTACCA GGCAGTGGGG CCCCAAGTGC

 901  AGGGCCATAG GAAGTACAAG TCTGGGAGAT ACTAGGCTGC ACTGTCTGTA

 951  GAGAATCTGA AAAAATAATA GAGTCACTGA AATGCAGTTT GGTATAATTA

1001  TTGCCATGCA TCATAATTCT AAATCATACT AGTGGTCAAA TACTCTTCCC

1051  TGAAAAAACA TTTTCTTGGT TTGAATTCTA AATAATTGTT GTGGTCACCA

1101  CTGAGCTTTT AAATATATAA ATACTTTCAA GTTTGCATAT TTTTATTACC
```

Fig. 9


```
1151  TGTTCCTTAA CAAACATTGA ATTCAACATG AAAATGATTA TGGGAAACAT
1201  TCGGGTATAC AGTCCCTGAC TCTTAAGGAC TCAGGTAAAT ACTTAGGGTA
1251  TTTCATGGCC CTAGTCTTTG GGGTACCACA TGTTTCTTCT TCAAATCACA
1301  GATTCAAAAT CAAGAATGAT AACACAGTGA TTGTGTAGAC AAAATAAGTG
1351  AACCAAAATT GCTTGCTTCT GTCATTCTAT GGAACCACTG AGAGTTTTTA
1401  CTTGTGCTTA AAATTTTGAA TAGTAAAACA GAGTGTCAAC TTCATGCTGG
1451  AATATTTTTG GCTTTTTAGA CACAATTTTA AGTACATGAA GTATTTTTAC
1501  AAGACTAAGT AACATCACTG AAATTACAGC TTTCTTCTTT TTAAAACTGG
1551  TATTTGTTAT AAAACTAAAG AGCGAATCAA GAAAAGCATA ATTATTACTG
1601  ATTATTACAG GATTATTACT GAAAAAGAAA TGTACGGAAT AGAGGAGGAA
1651  GGAGTTAACA AATGATCCAC TCTGGGTGTT GAAAACACCA ATAAGCCTGC
1701  TTCCAGGAAG TGCCTAAGAC AGAGCTGGCT CAGCTTGCTG GGTCACAGCA
1751  TGTAAGGAAA CTGCTGGGCT ACATGCCACC ATCCTCAGTT GTCCAGATAG
1801  ATAATCCCAT AGCCCCATGG GGAAATAATC TTTAATTATG ATATAGCTGA
1851  CACCATTCAA AGCACTATGC TAAGTCCTTT ATGTGAATTA ACTTTTGTCA
1901  AATTTATTTT TCATAAATAA CCCAAATATG TATACCACTA TTATCCTACC
1951  TTAAAGAGGA GAAACTGAGC TCCTAAAGTT TAAATATCTA ACCCAAGTTA
2001  AGACTGCTAG TCACCCTAGG CTATTAACTC AGGCAGTCTA ACTCAGGTAT
2051  AATAACATTA TGCTACTGTT TGCAGCTTTG ACTATGCCTG AATTATAACG
2101  TCATGCTATC TAACTAAAAA GCTAAGGGAA ATAAAATGAG CCATAGGGCT
2151  CAATTTCATA AAAGGAGAGA AAATACTGGG GAAAAGTGAT AATGCAGAGT
2201  TTAAAATATT TTTGTAAAAG TGCCAGAGAT TGAGTATAAC AAGTGTGACC
2251  AAAAAAAAAA AAAAAAAAAA AAAAGGAAGA AGGTAAAAAA AAGAGGGAGG
2301  TCTGAGAAAT AGAAATATCA GAGGAAGGAA ATAAAGGAGG GTGAGAGTAA
2351  ATTCTCTTTT AGCATTCAGA TTCCACAGAT TCCACAAATC ACATTTCTTT
2401  TTTTACCAAC TAAGGAAAAA TAACACTTGA CCTAACATTT CATTGCAGTT
```

Fig. 9


```
2451  AGCTAAAGGA TGCTAGAAAA ACTATGTTGC AGTGGTTTGC TCTAATTTCT

2501  TCAGGAATAG AGAAAAGTGA CAAAAAGATC AGAGAAGAGA AGAAAGGAAA

2551  CTATCAGAAA AATACAGAAT TGGAGTAGGA TATAACATAT TTGGGTTGAA

2601  GGTAAAATTT TATATTGTAA TCTTAAGTAT CTTGCTACTT CAGTTTGGTC

2651  CCTGGAACAG CAGCATCAGA ATCTGCCGAG GGCTTGTTAA AAAGGCAGAA

2701  TCTCAGGTCC CATCCCAGAC TCACTGAATC AGAATATAAA TACTGACAAG

2751  ATGCCCCGGG ATTCATATGC ACAGTAGAGC TGGCGAAGTT CCATTGTAGC

2801  CTGTGATTGT TTTCTGCAAC TTAGTATTTC TGAGTTTTCC CAAGGAAGAA

2851  AACCCAGGCC TTAGCTTCTG GCAGACTTGT GTTTCTCCTT TACTTACTAG

2901  CTGCATGACT CATGAGCAAG GAAATCAAAC TTTATGTGCC TGAGTTTCCT

2951  CATCTATAAA ATGGAGACTA TAATAATCAT CTCCTAGGCT TGTTTTGAGG

          M  F  N  K     C  S  F     H  S  S     I  Y  R  P     A  A  D
3001  ATGTTCAACA AATGCTCCTT TCATTCCTCT ATTTACAGAC CTGCCGCAGA

           N  S  A     S  S  L  C     A  I  I     C  F  L     N  L  V  I
3051  CAATTCTGCT AGCAGCCTTT GTGCTATTAT CTGTTTTCTA AACTTAGTAA

             E  C  D     L  E  T     N  S  E  I     N  K  L     I  I  Y
3101  TTGAGTGTGA TCTGGAGACT AACTCTGAAA TAAATAAGCT GATTATTTAT

          L  F  S  Q     N  N  R     I  R  F     S  K  L  L     L  K  I
3151  TTATTTTCTC AAAACAACAG AATACGATTT AGCAAATTAC TTCTTAAGAT

           L  F  Y     I  S  I  F     S  Y  P     E  L  M     C  E  Q  Y
3201  ATTATTTTAC ATTTCTATAT TCTCCTACCC TGAGTTGATG TGTGAGCAAT

             V  T  F     I  K  P     G  I  H  Y     G  Q  V     S  K  K
3251  ATGTCACTTT CATAAAGCCA GGTATACATT ATGGACAGGT AAGTAAAAAA

          H  I  I  Y     S  T  F     L  S  K     N  F  K  F     Q  L  L
3301  CATATTATTT ATTCTACGTT TTTGTCCAAA AATTTTAAAT TTCAACTGTT

           R  V  C     W  *
3351  GCGCGTGTGT TGGTAATGTA AAACAAACTC AGTACAGTAG TATTCAGTAC

3401  AGTATTTAAG CCCCTGTACT TAAACATATT CCTCGTACCA ATGAAGTTAC

3451  ATGAAAAGCA AATTTGTGTG AGATATCGTA GATGGAAGTA AATTAGTCTT
```

Fig. 9

3501 TATGTTCCCC ACAAATTGAA ATGCATTTCA AAAACTCTGT GTGTGTATGT

3551 GTGTGTGTGA CAGAGTGTGT GTGAGAGAGA GACAGAGAGA TACGCTTTGG

3601 TTGCCTCCAT AAGCTGGCTG CTATGATTAA TAAGACCAAG TTTTCTAAAG

3651 AAAATGAGAT CATAACAAAA GCCCTCTTTA TGACTATCTT TTATCAGGGG

3701 CAAAAAGGAA AGAGACAAAA CAGCATGAAA TGATGAGACC AAGTGATGAA

3751 AATTCATTCA CAATGATTGC TTTCAAGAGT AATTTCTCTT GGGTAATTCA

3801 GCAGCCTGTT ACTATGGCTC TCTGGAGTGA TAGCTAATGT AAATGAAGCC

3851 TCTAAAAGTG GATTATCCTG ACAAGAATAT ACTCAGCCAA TAATGCAACA

3901 GAAATCCATT CAAAGCATTC GGGAAAAATT CAAAAGAATA AATATTCTTT

3951 TTTTTTTTTT AAAGTTAATG ACCTACGATC CATTTCTTCC CTGACTAACA

4001 AGCAGCAAGC ACTTAAAAAT ATCCAGCCAG GATGAAATAG AAACCCACCT

4051 GACTTGTTAA TATTTTTGTT TGGTCCCAGG GACTCAGATT CTAAGCCAAA

4101 TTCTTTGAAT GATCTTGGCA AATGTCTCGA ATTATTTTTG CCAACTTTTC

4151 TTTATCTTGG AAAAAAAGTT TCATGAATGG GTGTCAAAAT TGATTAGTTT

4201 TAAAAACCTT TCTTGCAGAT ACGTATGGCA CCCTAAAACT GTATTAGAAA

4251 AAAATTTCAT ATCTCCAGGC CTTTCATTGG GTCAGGTTGG CATTTCGCTG

4301 CCCTTTATGT GTGTGACAAG TGAAAATAAG GAAAGAAAAA AACTCAAGTG

4351 AAGAAAATCA GAATCTGCGC AGCAGTTCCT GGGCGTTTCA GCTGCTTCCC

4401 ACATCACCTG CCTCATCAAG CCCCAGCATC CATCTCCTTG CTCATCTTAC

4451 ACCCTGTGTG CATGACAGGC CCACCATTCA TTTATCAGAG CAAAGGCTCT

4501 CCCACTATTC TGGTTCACCC CCCTACTTAG CCAGATATAC AAGAATATCT

4551 GCACGGATGA CCTGCCTCAC CTGGGAGCTC AGAGGAGCTC AGATTCCATT

4601 ACTATCGCAC CAAGGACAGA TCTCCCAGCA AGAATGACAG AAAAGACTAA

4651 CTGCCCCCAA AATCTCCCTT CCAAAACACA GTTCTCTTAA TTCTCCCAAG

4701 AAACCAGAAT GTGACTGCTC ACCTCTCTAA GGACCTGAAA ACAACTGGCC

4751 ATTTCAGCTA TTTAAATCAA CTTTAAAAAA TCCAACCGCC AAAATATTAA

Fig. 9


```
4801  ACCATTTTGG TTGGAATGAT AACATAACTA ACCTGCTGAC AGCTGCTTCT
4851  GCTAGGTGCA AAAATGGAAA AAAAAATACT TCTAATCAGG TCAAATCACT
4901  CTACCTTTGG GATTCTAAAT TTACTCATAT TCTCAAAGAA ATATATTCAG
4951  TCATAGTGGG GAAAATAGGA TTATTCCTTT AGCTCGATAA GCAACCAGAA
5001  GTTCTTCCTT CAAATCTTGA CATTTAATCA ATCAGAAATT GATTTTTGGA
5051  AAACTGTTTC CTATGAAGCT ATCTCTGCCT GAAGGATTTT TCTTTTACAA
5101  TCCAGACTAT AGAAGGAAAT TCACAACCTG GACTTTCACC TCCATTGGTC
5151  AGAGTTTTAC TGACCAATTC CCACCTCTGC CTTACACCTA ACGGAAGTTT
5201  ATGCCTGTTT TCTCTTCACA TACCCCAACA GTTACAAATG GTTGTTATTA
5251  TTAAGCATCT TTTATTTTGT GGCCTCTGAT TACATGGTCC CCTAAATTTT
5301  GACCTAATCA CAAAAGATTG GTAAAATTTC TTAACATATT AATAATATTT
5351  TGTTTATGTG TCAATATCTT AGCATGTATC AATTAAGACA GAGGTCTTAA
5401  CGTTCTCTTT TTGAAAGAGA ATATTAGGAT TCAGAGATAT TAAGAGATTC
5451  TCCCAGGATC ACAGTTAGGT AACAGAGCTG GATTTTAGTC CAGGTCTGTC
5501  TACAGCTCTA ACGTATATAC ACCCTTTGTA TAACATGTCA CGAATTCAGC
5551  ATAAAGGGAT CTTCAGTGAT CTAAGTCAGG GGTCAGCAAC CTTTTCTAAA
5601  AAGGACCAAA TAGTAATATT TCAGGCTTTG TGGACCCTAT GGTCTCTATC
5651  ATAACTGTTC AAATCACCAT GTAGTGTAAA AGGAGCCATA AGCAAAATAT
5701  AAACTAACGA ATGTGGCTGT TTTATGGGAT TTTTTTTTAA CTCTTTATTT
5751  ACAAAAGCAG GTGGCAGATC AGAACTCACT TATGGGCCAT AGTTCTCTGA
5801  CCCCTGACCT GAGAAAATCT TATATTTATG GACAACATTT AGACTGTGAC
5851  TTGCCAAGTA AGAACAAGAA GCTCTGTCAA CTGAAGGTCA AGGCTGGAGT
5901  TCTGAAAGCA AAGAGCTGTC TGGTGTTAAT GATAAGTGAA ATAGTTAAAG
5951  TTAGAAGATC CCAGTTATAA GAAGCACAAA GAATAATGAC CATAGACTCC
6001  TGAACAAGAA TGTCTGGACT TCTGGCTTAG GCACTCTTGT TGTATGGTCC
```

Fig. 9


```
6051  AGGCCAAGTT ACCTAATCTC TCCAGGCCTC CATTTTCTTA TCATTAAATG
6101  AAGATAATAA AAGTATTTTC CTCAGAGAGC TGTAAGAATA AACTGAGCTA
6151  ACCCATGTCA AGCACATAGA ATAGGGCCCA GCCTATATTA ATTTATCAAT
6201  AAATGCCAG
```

Fig. 10.

Sequence of the AMB1 protein

```
M  F  N  K  C  S  F  H  S  S  I  Y  R  P  A  A  D
N  S  A  S  S  L  C  A  I  I  C  F  L  N  L  V  I
E  C  D  L  E  T  N  S  E  I  N  K  L  I  I  Y
L  F  S  Q  N  N  R  I  R  F  S  K  L  L  L  K  I
L  F  Y  I  S  I  F  S  Y  P  E  L  M  C  E  Q  Y
V  T  F  I  K  P  G  I  H  Y  G  Q  V  S  K  K
H  I  I  Y  S  T  F  L  S  K  N  F  K  F  Q  L  L
R  V  C  W  *
```

Fig. 11

	1	2	3	4	5	6	7	8	9	10	11	12
A	whole brain	cerebellum left	substantia nigra	heart	esophagus	colon, transverse	kidney	lung	liver	leukemia, HL-60	fetal brain	yeast total RNA
B	cerebral cortex	cerebellum right	accumbens nucleus	aorta	stomach	colon, desending	skeletal muscle	placenta	pancreas	HeLa S3	fetal heart	yeast tRNA
C	frontal lobe	corpus callosum	thalamus	atrium, left	duodenum	rectum	spleen	bladder	adrenal gland	leukemia K-562	fetal kidney	*E.coli* rRNA
D	parietal lobe	amygdala	pituitary gland	atrium, right	jejunum		thymus	uterus	thyroid gland	leukemia, MOLT-4	fetal liver	*E.coli* DNA
E	occipital lobe	caudate nucleus	spinal cord	ventricle left	ileum		peripheral blood leukocyte	prostate	salivary gland	Burkitt's lymphoma, Raji	fetal spleen	Poly r(A)
F	temporal lobe	hippo-campus		ventricle right	ilocecum		lymph node	testis	mammary gland	Burkitt's lymphoma, Daudi	fetal thymus	human Cot-1 DNA
G	p.g.* of cerebral cortex	medulla oblongata		inter-ventricular septum	appendix		bone marrow	ovary		colorectal adeno-carcinoma SW480	fetal lung	human DNA 100 ng
H	pons	putamen		apex of the heart	colon, ascending		trachea			lung carcinoma A549		human DNA 500 ng